A N-BODY SOLVER FOR SQUARE ROOT ITERATION *

MATT CHALLACOMBE*†, TERRY HAUT*AND NICOLAS BOCK*

Abstract. We develop the Sparse Approximate Matrix Multiply (`SpAMM`) n-body solver for first order Newton Schulz iteration of the matrix square root and inverse square root. The solver performs recursive two-sided metric queries on a modified Cauchy-Schwarz criterion, culling negligible sub-volumes of the product-tensor for problems with structured decay in the sub-space metric. These sub-structures are shown to bound the relative error in the matrix-matrix product, and in favorable cases, to enjoy a reduced computational complexity governed by dimensionality reduction of the product volume. A main contribution is demonstration of a new, algebraic locality that develops under contractive identity iteration, with collapse of the metric-subspace onto the identity's plane diagonal, resulting in a stronger `SpAMM` bound. Also, we carry out a first order Fréchet analyses for single and dual channel instances of the square root iteration, and look at bifurcations due to ill-conditioning and a too aggressive `SpAMM` approximation. Then, we show that extreme `SpAMM` approximation and contractive identity iteration can be achieved for ill-conditioned systems through regularization, and we demonstrate the potential for acceleration with a scoping, product representation of the inverse factor.

1. Introduction. In many areas of current numerical interest, matrix equations with decay properties describe correlations over a range of scales. By decay, we mean an approximate inverse relationship between a matrix element's magnitude and an associated distance; this might be a slow inverse exponential relationship between matrix elements and a Cartesian separation, or it might involve a non-Euclidean distance, *e.g.* between character strings.

A common approach to exploiting matrix decay involves sparse approximation of inverse factors that transform Gramian equations to a representation independent form, via congruence transformations based on Löwdin's symmetric orthogonalization (the matrix inverse square root) [85, 91], inverse Cholesky factorization [75] or related transformations that involve an inverse or pseudo-inverse [61, 18, 55, 56]. Gramian inverse factors with decay are ubiquitous to problems with local, non-orthogonal support, including finite element calculations [40, 59], radial-basis-function finite-difference calculations [114, 110], in the "direct" approach to radial-basis interpolation [106], with frames [47, 63], with computation involving "lets" of various types [55, 56], and in the Atomic Orbital (AO) representation [71, 64].

Off-diagonal decay of the matrix sign function is also a well developed area of study in statistics and statistical physics [98, 117, 6, 60, 76], and in electronic structure, where sparse approximation enables fast computation of the the gap shifted matrix sign function as projector of an effective Hamiltonian [17, 13, 22, 14]. Short to long ranged decay properties of the projector are shown in Fig. 1.1. These matrix functions, the matrix sign function and the matrix inverse square root, are related by Higham's identity [66]:

$$\operatorname{sign}\left(\begin{bmatrix} 0 & \boldsymbol{s} \\ \boldsymbol{I} & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & \boldsymbol{s}^{1/2} \\ \boldsymbol{s}^{-1/2} & 0 \end{bmatrix}. \tag{1.1}$$

A well conditioned matrix $\boldsymbol{s}$ may often correspond to matrix sign and inverse square root functions with rapid exponential decay, and be amenable to *ad hoc*

*This article was released under LA-UR-15-26304. The Los Alamos National Laboratory is operated by Los Alamos National Security, LLC for the NNSA of the USDoE under Contract No. DE-AC52- 06NA25396.

*Los Alamos National Laboratory

†matt.challacombe@freeon.org

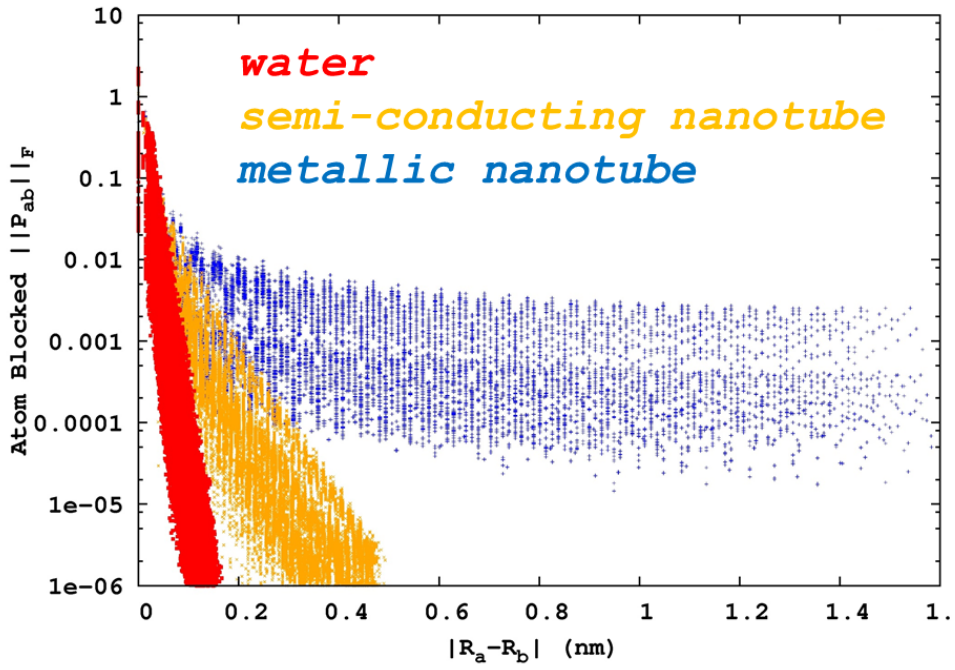

FIG. 1.1. *Examples from electronic structure of decay for the spectral projector (gap shifted sign function) with respect to the local (atomic) support. Shown is decay for systems with correlations that are short (insulating water), medium (semi-conducting 4,3 nanotube), and long (metallic 3,3 nanotube) ranged, from exponential (insulating) to algebraic (metallic).*

matrix truncation or "sparsification", $\bar{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{\epsilon}_\tau^{\boldsymbol{s}}$, where $\boldsymbol{\epsilon}_\tau^{\boldsymbol{s}}$ is the error introduced according to some criterion τ, supported by useful bounds to matrix function elements [15, 16, 97, 59, 26]. The criterion τ might be a drop-tolerance, $\epsilon_\tau^{\boldsymbol{s}} = \{-s_{ij} * \hat{\boldsymbol{e}}_i \,|\, |s_{ij}| < \tau\}$, a radial cutoff, $\epsilon_\tau^{\boldsymbol{s}} = \{-s_{ij} * \hat{\boldsymbol{e}}_i \,|\, \|\boldsymbol{r}_i - \boldsymbol{r}_j\| > \tau\}$, or some other approach to truncation, perhaps involving a sparsity pattern chosen *a priori* for computational expedience. Then, the sparse general matrix-matrix multiply (`SpGEMM`) [58, 113, 28, 21] may be employed, yielding fast solutions for multiplication rich iterations, and with fill-in modulated by truncation. Exhaustive surveys of these methods in the numerical linear algebra are given by Benzi [17, 13], and by Bowler [22] and Benzi [14] for electronic structure.

In addition to sparsity, data localities leading to high operation counts are essential for kernels like the `SpGEMM` and their distributed implementations. Over the past decades, methods have evolved from bandwidth reduction (Cuthill-McKee) + greedy blocking [113], progressing with tours of the graph via heuristic solutions to the Traveling Salesman Problem (TSP) [100, 3, 81], and more recently towards reordering based on cache modeling and dynamic sampling [48, 99]. Ordering with graph partitioning, targeting the load balance, may also lead to exploitable localities, via *e.g.* proximity to the diagonal [24]. Of current interest are ordering schemes that enhance the weighted block-locality of the Page Rank problem [73, 39, 82, 128].

Matrix locality may also result from an ordering that preserves locality in an auxiliary representation, a property of sub-space mappings that preserve local neighborhoods [10, 11, 12]. In the case of electronic structure, Space Filling Curve (SFC)

heuristics applied to a local Cartesian basis results in Gramian matrices with neighborhoods segregated by magnitude [28, 23], as shown in Fig. (1.2). Likewise, Sierpinski curves and Self Avoiding Walks on meshes lead to locality preserving orderings [62, 7], for *e.g.* finite elements [93, 108]. This type of weighted block-locality or "Block-By-Magnitude" (BBM) structure of the subspace metric $\|\cdot\|_F$ is finely resolved with the quadtree matrix [125, 1, 101, 124, 126, 84];

$$\boldsymbol{a}^i = \begin{bmatrix} \boldsymbol{a}_{00}^{i+1} & \boldsymbol{a}_{01}^{i+1} \\ \boldsymbol{a}_{10}^{i+1} & \boldsymbol{a}_{11}^{i+1} \end{bmatrix}, \tag{1.2}$$

where i is the recursion depth, and

$$\|\boldsymbol{a}^i\|_F = \sqrt{\|\boldsymbol{a}_{00}^{i+1}\|_F^2 + \|\boldsymbol{a}_{01}^{i+1}\|_F^2 + \|\boldsymbol{a}_{10}^{i+1}\|_F^2 + \|\boldsymbol{a}_{11}^{i+1}\|_F^2}\,, \tag{1.3}$$

is the sub-multiplicative Frobenius norm [51, 65, 72].

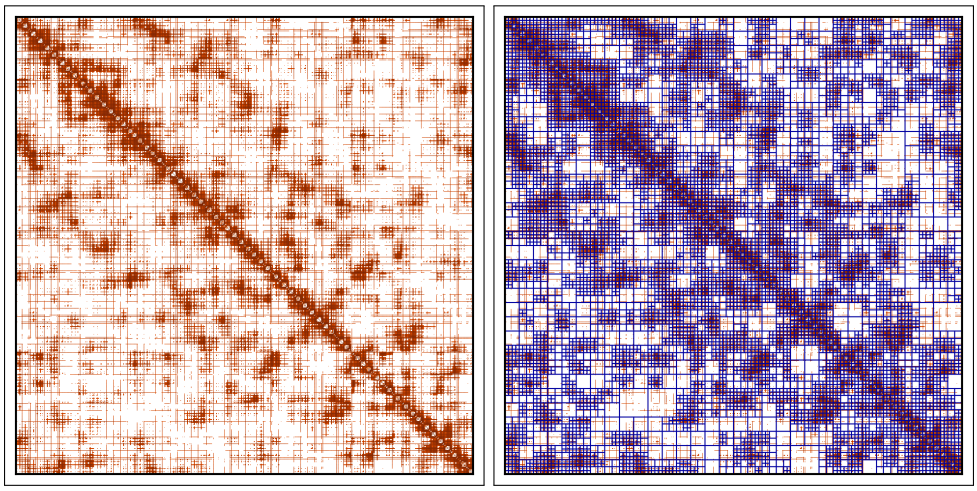

FIG. 1.2. *At left, Block-By-Magnitude (BBM) structure of a quantum chemical Gramian (overlap matrix), for a box of 100 water molecules, with Cartesian support along a locality preserving curve. At right, quadtree resolution of neighborhoods with norms down to* 10^{-3}.

However, despite structuring for cache, distributed memory or to enhance BBM structuring, matrix truncation may still be ineffective for ill-conditioned problems, because the rate of decay may be too slow, and also because of increased numerical sensitivities to the sparse approximation:

$$\overline{\boldsymbol{a}\cdot\boldsymbol{b}} \;=\; \boldsymbol{a}\cdot\boldsymbol{b} \;+\; \boldsymbol{\epsilon}_\tau^{\boldsymbol{a}}\cdot\boldsymbol{b} \;+\; \boldsymbol{a}\cdot\boldsymbol{\epsilon}_\tau^{\boldsymbol{b}} \;+\; \mathcal{O}(\tau^2)\,, \tag{1.4}$$

allowing to control only absolute errors. An alternative approach is to find a reduced rank approximation, ideally closed under the operations of interest. However, rank reduction may be expensive if the rank is not much, much smaller than the dimension. Interestingly, in the ultra-flat limit, kernel methods enjoy rank reduction corresponding formally to change of basis, enabling fast methods for constructing the generalized inverse [107, 31]. In cases with simply slow exponential decay however, our experience has so far been that naive element dropping is about as effective as dropping singular values.

In this contribution, we consider the regime between trivial sparsity and formal rank reduction, with fast multiplications exploiting instead an accelerated *volumetric decay* in subspace-metric of the product-tensor. There, the `SpAMM` kernel carries out octree scoping of low-dimensional structures that bound the relative error while yielding a reduced complexity multiplication. Beyond decay associated with the the matrix square root and its inverse, we demonstrate additional compression of these bounding volumes under contractive identity iteration.

This paper is organized as follows: In Section 2, we modify the `SpAMM` occlusion-cull to bound the relative product error. In Section 3, we review several instances of the first order Newton-Schulz (NS) square root iteration, and go over the contractive identity iteration that develops in the basin of stability. In Section 4, we overview a generic implementation of the `SpAMM` kernel and introduce quantum chemical and engineering data of interest. In Section 5, we develop a Fréchet analysis for NS instances and the `SpAMM` algebra, and examine error flows in bifurcating and stable square root iterations for ill-conditioned problems. In Section 6 we show that even difficult, ill-conditioned problems can be brought to the regime of strongly contractive identity iteration, through iterative regularization and precision scoping. In Section 7, we show for the first time the process of lensing, involving sub-space contraction to diagonal planes of the ijk-cube ($i = j$, $i = k$ and/or $j = k$), followed finally by compression onto the identity's plane diagonal, yielding additional orders of magnitude compression of `SpAMM` sub-volumes. Finally, in Section 8 we argue it may be possible to remain close to the lensed state whilst constructing a deferred product representation of the inverse factor.

2. `SpAMM`**.** The Sparse Approximate Matrix Multiply (`SpAMM`) is a reduced complexity approximation that evolved from a row-column skipout mechanism within the blocked-compressed-sparse-row (BCSR) [27] and the distributed-blocked-compressed-row (DBCSR) data structures [29], to methods with fast subspace resolution through octree recursion [32, 19, 20]. Finding sub-spaces via fast range or metric query is a generic n-body problem handled with agility by the quadtree [77, 50, 102, 52], a problem related to spatial hashing [112, 80] and the occlusion-cull in visualization [96].

The `SpAMM` kernel $\otimes_\tau$ provides fast approximate multiplication for matrices with decay and metric locality, with errors controlled by the scoping parameter τ:

$$\widetilde{\boldsymbol{a}\cdot\boldsymbol{b}} \equiv \boldsymbol{a}\otimes_\tau\boldsymbol{b} = \boldsymbol{a}\cdot\boldsymbol{b} + \boldsymbol{\Delta}_\tau^{a\cdot b}\,. \tag{2.1}$$

As $\tau \to 0$, `SpAMM` reverts to the recursive `GEMM` [57, 46].

In this work, we promote the following stable version of the `SpAMM` occlusion-cull:

$$\boldsymbol{a}^i\otimes_\tau\boldsymbol{b}^i = \begin{cases} \mathbf{0} & \texttt{if} \quad \|\boldsymbol{a}^{\mathtt{i}}\|\|\boldsymbol{b}^{\mathtt{i}}\| < \tau\|\mathtt{a}\|\|\mathtt{b}\| \\ \boldsymbol{a}^i\cdot\boldsymbol{b}^i & \texttt{if}(\mathtt{i}=\mathtt{leaf}) \\ \begin{bmatrix} \boldsymbol{a}_{00}^{i+1}\otimes_\tau\boldsymbol{b}_{00}^{i+1} + \boldsymbol{a}_{01}^{i+1}\otimes_\tau\boldsymbol{b}_{10}^{i+1}\,, & \boldsymbol{a}_{00}^{i+1}\otimes_\tau\boldsymbol{b}_{01}^{i+1} + \boldsymbol{a}_{01}^{i+1}\otimes_\tau\boldsymbol{b}_{11}^{i+1} \\ \boldsymbol{a}_{00}^{i+1}\otimes_\tau\boldsymbol{b}_{01}^{i+1} + \boldsymbol{a}_{01}^{i+1}\otimes_\tau\boldsymbol{b}_{11}^{i+1}\,, & \boldsymbol{a}_{00}^{i+1}\otimes_\tau\boldsymbol{b}_{01}^{i+1} + \boldsymbol{a}_{01}^{i+1}\otimes_\tau\boldsymbol{b}_{11}^{i+1} \end{bmatrix} & \texttt{else} \end{cases}, \tag{2.2}$$

with $\|\cdot\| \equiv \|\cdot\|_F$ and the leaf condition determined by the block size, N_b. This scoping partitions the product tensor into two sub-spaces: the space of culled leaf-tasks, $\boldsymbol{a}\otimes_\tau\boldsymbol{b}$, and its complement, the occlusion error $\boldsymbol{\Delta}_\tau^{a\cdot b}$ of avoided multiplications. This occlusion

error is bounded by

$$\frac{\|\boldsymbol{\Delta}_\tau^{a\cdot b}\|}{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|} \le n^2\tau\,, \tag{2.3}$$

as shown in the following section, a result commensurate with the stable, normwise multiplication criteria emphasized by Demel, Dumitriu, Holtz and Kleinberg (DDHK) [42].

2.1. Bound. We now prove Eq. (2.3):

PROPOSITION 2.1. *Let* $\tau_{ab} = \tau\|\boldsymbol{a}\|\|\boldsymbol{b}\|$. *Then for each* i, j,

$$\left|(\boldsymbol{a}\otimes_\tau\boldsymbol{b})_{ij} - (\boldsymbol{a}\cdot\boldsymbol{b})_{ij}\right| \le n\,\tau_{ab},$$

and

$$\|\boldsymbol{a}\otimes_\tau\boldsymbol{b} - \boldsymbol{a}\cdot\boldsymbol{b}\| \le n^2\,\tau_{ab}.$$

Proof. We first show the following technical result: it is possible to choose $\alpha_{lij} \in \{0,1\}$ such that

$$(\boldsymbol{a}\otimes_\tau\boldsymbol{b})_{ij} = \sum_{l=1}^{n} a_{il}\,b_{lj}\,\alpha_{lij}, \tag{2.4}$$

In addition, if $\alpha_{lij} = 0$, then $|a_{il}|\,|b_{lj}| < \tau_{ab}$. To show this, we use induction on the number $k_{\max}$ of levels.

First, if $k_{\max} = 0$,

$$\boldsymbol{a}\otimes_\tau\boldsymbol{b} = \begin{cases} 0 & \text{if } \|\boldsymbol{a}\|\,\|\boldsymbol{b}\| < \tau_{ab}, \\ \boldsymbol{a}\cdot\boldsymbol{b} & \text{else.} \end{cases}$$

Therefore, $\boldsymbol{a}\otimes_\tau\boldsymbol{b}$ is of the form (2.4) with either all $\alpha_{lij} = 0$ or all $\alpha_{lij} = 1$. Moreover, if $\alpha_{lij} = 0$, then $|a_{il}|\,|b_{lj}| \le \|\boldsymbol{a}\|\,\|\boldsymbol{b}\| < \tau_{ab}$.

Now assume that the claim holds for $k_{\max} - 1$. We show that it holds for $k_{\max}$. Indeed, if $\|\boldsymbol{a}\|\,\|\boldsymbol{b}\| < \tau_{ab}$, we have that $\boldsymbol{a}\otimes_\tau\boldsymbol{b} = 0$, which is of the form (2.4) with all $\alpha_{lij} = 0$. Also, if $\alpha_{lij} = 0$, then $|a_{il}|\,|b_{lj}| < \|\boldsymbol{a}\|\,\|\boldsymbol{b}\| < \tau_{ab}$.

Now assume that $\|\boldsymbol{a}\|\,\|\boldsymbol{b}\| \ge \tau_{ab}$. Then

$$\boldsymbol{a}\otimes_\tau\boldsymbol{b} = \begin{pmatrix} \boldsymbol{a}_{00}\otimes_\tau\boldsymbol{b}_{00} + \boldsymbol{a}_{01}\otimes_\tau\boldsymbol{b}_{10} & \boldsymbol{a}_{00}\otimes_\tau\boldsymbol{b}_{01} + \boldsymbol{a}_{01}\otimes_\tau\boldsymbol{b}_{11} \\ \boldsymbol{a}_{10}\otimes_\tau\boldsymbol{b}_{00} + \boldsymbol{a}_{11}\otimes_\tau\boldsymbol{b}_{10} & \boldsymbol{a}_{10}\otimes_\tau\boldsymbol{b}_{10} + \boldsymbol{a}_{11}\otimes_\tau\boldsymbol{b}_{11} \end{pmatrix}.$$

We need to consider four cases: $i \le n/2$ and $j \le n/2$, $i > n/2$ and $j > n/2$, $i > n/2$ and $j \le n/2$, and, finally, $i > n/2$ and $j > n/2$. Since the analysis is similar for all

four cases, we only consider $i \leq n/2$ and $j \leq n/2$. We have that

$$\begin{aligned}(\boldsymbol{a}\otimes_\tau \boldsymbol{b})_{ij} &= (\boldsymbol{a}_{00}\otimes_\tau \boldsymbol{b}_{00} + \boldsymbol{a}_{01}\otimes_\tau \boldsymbol{b}_{10})_{ij} \\ &= \sum_{l=1}^{n/2} (\boldsymbol{a}_{00})_{il} (\boldsymbol{b}_{00})_{lj} \alpha^0_{lij} \\ &\qquad + \sum_{l=1}^{n/2} (\boldsymbol{a}_{01})_{il} (\boldsymbol{b}_{10})_{lj} \alpha^1_{lij} \\ &= \sum_{l=1}^{n} a_{il} b_{lj} \alpha_{lij},\end{aligned}$$

where we used the induction hypothesis in the second equality.

Now suppose that $\alpha_{lij} = 0$ for some l. Then $\tilde{\alpha}^0_{lij} = 0$ if $l \leq n/2$ or $\tilde{\alpha}^1_{l-n/2,ij} = 0$ if $l > n/2$. If, e.g., $\tilde{\alpha}^1_{l-n/2,ij} = 0$, then $|a_{il}|\,|b_{lj}| = \left|(\boldsymbol{a}_{01})_{i,l-n/2}\right| \left|(\boldsymbol{b}_{10})_{l-n/2,j}\right| < \tau_{ab}$, where we used the induction hypothesis in the final inequality. The analysis for $l \leq n/2$ is similar, and the claim follows.

We can now finish the proof of Proposition 2.1. Indeed, by (2.4),

$$\begin{aligned}\left|(\boldsymbol{a}\otimes_\tau \boldsymbol{b})_{ij} - (\boldsymbol{a}\cdot\boldsymbol{b})_{ij}\right| &\leq \sum_{l=1}^{n} |a_{il} b_{lj}|\,|\alpha_{lij} - 1| \\ &= \sum_{\alpha_{lij}=0} |a_{il} b_{lj}|\,.\end{aligned}$$

In addition, if $\alpha_{lij} = 0$, then $|a_{il} b_{lj}| < \tau_{ab}$ and the lemma follows.
□

2.2. Related research. `SpAMM` is perhaps most closely related to the Strassen-like branch of fast matrix multiplication [111, 41, 9, 79, 5], and also methods for group theoretical embedding allowing fast polynomial multiplication [38, 37, 115]. In the Strassen-like approach, disjoint volumes in high order tensor expansions of the product are recursively excluded, while in the `SpAMM` approach to fast multiplication, the subspace metric of the product tensor is recursively queried for occlusion of negligible volumes, with error bounded by Eq. (2.3). These methods for fast matrix multiplication are stable, satisfying the DDHK normwise product bound [41].

This work offers a data local alternative to fast non-deterministic methods for sampling the product, which include sketching [105, 44, 86, 94, 127], joining [90, 68, 70, 34, 4, 83, 74], sensing [69] and probing [36]. These methods may involve probabilistic assumptions and on the fly sampling, with the potential for complexity reduction due to statistical approximations. `SpAMM` also employs on the fly weighted sampling, with compression through octree scoping of metric tensor decay, and with additional subspace compression due to the onset of identity iteration.

`SpAMM` is related to the generalized n-body methods popularized by Gray [53, 54]. Here and in related research, we are interested in generic approaches to approximation that are data agnostic, based on the quadtree and its generalizations [77, 50, 102, 52] and and on the facile measure $\|\cdot\| \equiv \|\cdot\|_F$ [72]. In this work, a fast two-sided metric query enables octree scoping with the occlusion criteria $\|\boldsymbol{a}^i\|\|\boldsymbol{b}^i\| < \tau\|a\|\|b\|$. With quantum chemical Fock exchange, a fast three-sided metric query enables hextree scoping with a related, Cauchy-Schwarz like occlusion criterion (direct SCF) [30]. It

may also be possible to exploit subspace locality more broadly, though mappings that optimally preserve local neighborhoods in higher dimensions, *e.g.* via the Laplace-Beltrami operator [10, 11, 12].

For distributed architectures, n-body methods offer well established protocols for turning spatial locality into data and temporal locality [119, 120, 121, 122, 118]. Recently, we showed strong scaling for the `SpAMM` kernel [20], while Driscoll *et. al* were able to show perfect strong scaling and communication optimally for pairwise n-body methods [45]. A uniform approach to generic scoping is empowered at the ecosystems level by runtime support for recursive task parallelism [43, 103, 78, 88, 104, 89, 87].

Finally, this work is inspired broadly by Higham's work, particularly by Higham, Mackey, Mackey and Tisseur (HMMT) in 2005 [67] on square root iteration and the group structure of matrix functions. Also, it is influenced by Chen and Chow's [33] approach to scaled NS iteration for ill-conditioned problems, and by the Helgaker group's work on NS iteration, whose notation we follow in part [71].

3. Newton-Shulz Iterations. There are two common, first order NS iterations; the sign iteration and the square root iteration, related by the square $\boldsymbol{I}(\cdot) = \text{sign}^2(\cdot)$. These equivalent iterations converge linearly at first, then enter a basin of stability marked by super-linear convergence.

3.1. Sign iteration. For the NS sign iteration, this basin is marked by a behavioral change in the difference $\delta \boldsymbol{X}_k = \widetilde{\boldsymbol{X}}_k - \boldsymbol{X}_k = \text{sign}(\boldsymbol{X}_{k-1} + \delta \boldsymbol{X}_{k-1}) - \text{sign}(\boldsymbol{X}_{k-1})$, where $\delta \boldsymbol{X}_{k-1}$ is some previous error. The change in behavior is associated with the onset of idempotence and the bounded eigenvalues of $\text{sign}'(\cdot)$, leading to stable iteration when $\text{sign}'(\boldsymbol{X}_{k-1})\,\delta \boldsymbol{X}_{k-1} < 1$. Global perturbative bounds on this iteration have been derived by Bai and Demmel [8], while Byers, He and Mehrmann [25] developed asymptotic bounds. The automatic stability of sign iteration is a well developed theme in Higham's work [66].

3.2. Square root iteration. We are concerned with resolution of the identity

$$\boldsymbol{I}(\boldsymbol{s}) = \boldsymbol{s}^{1/2} \cdot \boldsymbol{s}^{-1/2}, \tag{3.1}$$

and its low-complexity computation with fast methods.

Starting with eigenvalues rescaled to the domain $(0, 1]$ with the easily obtained largest eigenvalue, $\boldsymbol{s} \leftarrow \boldsymbol{s}/s_{N-1}$, and with $\boldsymbol{z}_0 = \boldsymbol{I}$ and $\boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{s}$, the corresponding canonical, "dual" channel square root iteration is:

$$\begin{aligned} \boldsymbol{y}_k &\leftarrow h_\alpha\left[\boldsymbol{y}_{k-1} \cdot \boldsymbol{z}_{k-1}\right] \cdot \boldsymbol{y}_{k-1} \\ \boldsymbol{z}_k &\leftarrow \boldsymbol{z}_{k-1} \cdot h_\alpha\left[\boldsymbol{y}_{k-1} \cdot \boldsymbol{z}_{k-1}\right] \ , \end{aligned} \tag{3.2}$$

converging as $\boldsymbol{y}_k \to \boldsymbol{s}^{1/2}$, $\boldsymbol{z}_k \to \boldsymbol{s}^{-1/2}$ and $\boldsymbol{x}_k \to \boldsymbol{I}$, with eigenvalues aggregated towards 1 by the NS map $h_\alpha[\boldsymbol{x}] = \frac{\sqrt{\alpha}}{2}(3 - \alpha \boldsymbol{x})$ [66, 67]. As in the case of sign iteration, this canonical iteration was shown by Higham, Mackey, Mackey and Tisseur [67] to remain strongly bounded in the super-linear regime, by idempotent Fréchet derivatives about the fixed point $\left(\boldsymbol{s}^{1/2}, \boldsymbol{s}^{-1/2}\right)$, in the direction $\left(\delta \boldsymbol{y}_{k-1}, \delta \boldsymbol{z}_{k-1}\right)$:

$$\delta \boldsymbol{y}_k = \frac{1}{2}\delta \boldsymbol{y}_{k-1} - \frac{1}{2}\boldsymbol{s}^{1/2} \cdot \delta \boldsymbol{z}_{k-1} \cdot \boldsymbol{s}^{1/2} \tag{3.3}$$

$$\delta \boldsymbol{z}_k = \frac{1}{2}\delta \boldsymbol{z}_{k-1} - \frac{1}{2}\boldsymbol{s}^{-1/2} \cdot \delta \boldsymbol{y}_{k-1} \cdot \boldsymbol{s}^{-1/2}\ . \tag{3.4}$$

In addition to the dual channel instance, we also consider the "single" channel version of square root iteration,

$$\begin{aligned} \boldsymbol{z}_k &\leftarrow \boldsymbol{z}_{k-1} \cdot h_\alpha \left[\boldsymbol{x}_{k-1} \right] \ , \\ \boldsymbol{x}_k &\leftarrow \boldsymbol{z}_k^T \cdot \boldsymbol{s} \cdot \boldsymbol{z}_k \ . \end{aligned} \tag{3.5}$$

4. Implementation.

4.1. Programming. In our experimental research, issue driven implementations of `SpAMM` have been developed, including a Haskell version (formal functional programming) [35], a fine grained (4×4) `single-precision` assembly coded version (scalar performance) [19] and a task parallel version in `C++`, `OpenMP 3.0` and `Charm++` (strong scaling) [20]. In the current contribution, informal functional programming in `Fortran08` was used, with the goal of generic simplicity and mathematical agility.

In our implementation, allocation functions instantiate or reuse sub-matrices in downward recursion, and accumulate decorations (flops, bounding boxes, non-zeros, norms, initialization flags *etc.*) in backwards recursion, up the stack. Optional, `ifdef`'d features include the first order Fréchet analyses, outlined in Section 5.1 and using `MATMUL`, as well as sparse `VTK` output for visualization of the ijk product volumes, shown in Section 7.

Precision is determined by the block dimension N_b, the primary threshold τ controlling error in the $\boldsymbol{z}$ and the $\boldsymbol{x}$ channels, and by the tighter (sensitive) threshold τ_s for the $\boldsymbol{y}$ channel. Unless stated otherwise, we take $N_b = 16$ and $\tau_s \sim .01 \times \tau$. Finally, reported calculations were carried out in double precision using the `GCC/gfortran 4.8.1` compiler.

4.2. Mapping. The NS logistic map for the square root iteration is $h_\alpha[\boldsymbol{x}] = \frac{\sqrt{\alpha}}{2}(3 - \alpha\boldsymbol{x})$, with the initial rate of convergence controlled by h'_α and the smallest eigenvalue, x_0. Various schemes for controlling the values α towards convergence include methods by Pan and Schreiber [95], and more recently, Jie and Chen [33], who demonstrated $2\times$ acceleration for very ill-conditioned problems with their continuous scaling approach.

In addition to scaling of the NS logistic, we introduce a stabilizing map that accounts for eigenvalues tossed out of bounds by $\otimes_\tau$. This stabilization is the transformation $[0,1] \to [0+\varepsilon, 1-\varepsilon]$ (shift and scale), carried out prior to application of the logistic.

The most important aspect of these scaling and stabilization maps is to turn them off towards convergence. Conventional methods often compute a lowest eigenvalue to monitor convergence [95, 33], but this may be too expensive for ill-conditioned problems. Alternatively, we monitor convergence simply with the relative trace error, $t_k = (n - \operatorname{tr} \widetilde{\boldsymbol{x}}_k)\, n^{-1}$. Then, sigmoidal functions damp scaling to unity,

$$\alpha(t) = 1. + 1.85 \times \left(1 + e^{-50.(t-.35)} \right)^{-1} , \tag{4.1}$$

and the stability parameter to zero,

$$\varepsilon(t) = .1 \times \left(1 + e^{-75.(t-.30)} \right)^{-1} . \tag{4.2}$$

These empirical damping functions are used throughout.

4.3. Data. Data for numerical experiments include problems from electronic structure and structural engineering. Electronic structure matrices were obtained from the non-orthogonal metric (overlap matrix) of the generalized eigenproblem, encountered in local support with Gaussian-Type Atomic-Orbitals (GTAOs) [64]. A sequence of nanotube matrices, $36\times \to 128\times$ the (3,3) unit cell, were generated with `TubeGen` [49] and a modified STO-2G [109] basis, with an added diffuse (flat) Gaussian *sp*-shell and exponents $\zeta_{10^{10}} =$ `.06918` and $\zeta_{10^{11}} =$ `.05934`, corresponding to the condition numbers $\kappa(\boldsymbol{s}) = 10^{10}$ and $\kappa(\boldsymbol{s}) = 10^{11}$ respectively[1]. We also constructed a sequence of matrices from periodic water cubes, in increments of 100, with coordinates obtained using the `gromacs` utility `gmx solvate -box` [116] and the triple-ζ 6-311G** GTAO basis [109]. While less ill-conditioned than the nano-tube sequence, $\kappa(\boldsymbol{s}) \sim 10^5$, the water cube matrices manifest a different metric locality due to dimensionality. Also, we experiment with the `bcsstk14` structural engineering matrix for the Roof of the Omni Coliseum [123].

5. Error Flow.

5.1. Stability. Stability in the square root iteration is determined by the differential

$$\delta \boldsymbol{x}_k = \boldsymbol{x}_{\delta \widehat{\boldsymbol{y}}_{k-1}} \times \delta y_{k-1} + \boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} \times \delta z_{k-1} + \mathcal{O}(\tau^2)\,, \tag{5.1}$$

which must remain bounded below one to avoid divergence. The corresponding Fréchet derivatives are

$$\boldsymbol{x}_{\delta \widehat{\boldsymbol{y}}_{k-1}} = \lim_{\tau \to 0} \frac{\boldsymbol{x}\left(\boldsymbol{y}_{k-1} + \tau \delta \widehat{\boldsymbol{y}}_{k-1}, \boldsymbol{z}_{k-1}\right) - \boldsymbol{x}_k}{\tau} \tag{5.2}$$

and

$$\boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} = \lim_{\tau \to 0} \frac{\boldsymbol{x}\left(\boldsymbol{y}_{k-1}, \boldsymbol{z}_{k-1} + \tau \delta \widehat{\boldsymbol{z}}_{k-1}\right) - \boldsymbol{x}_k}{\tau}\,, \tag{5.3}$$

along unit directions of the previous errors $\delta \widehat{\boldsymbol{y}}_{k-1}$ and $\delta \widehat{\boldsymbol{z}}_{k-1}$, by an amount determined by the displacements $\delta y_{k-1} = \|\delta \boldsymbol{y}_{k-1}\|$ and $\delta z_{k-1} = \|\delta \boldsymbol{z}_{k-1}\|$. In the single instance, we have simply:

$$\delta \boldsymbol{x}_k = \boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} \times \delta z_{k-1} + \mathcal{O}(\tau^2)\,. \tag{5.4}$$

This formulation makes plain changes about the resolvent, separating orientational effects for derivatives of the unit direction, set mostly by the underlying exact linear algebra, from changes to error displacements, which involve both the action of derivatives on previous errors, as well as current `SpAMM` occlusion errors. In the following sections, we develop this form of the error. Then, in Section 5.4, we show interesting behaviors of these derivatives at the edge of stability.

5.2. Fréchet derivatives. In the dual instance, Fréchet derivatives occurring in Eq. (5.1) are:

$$\begin{aligned}\boldsymbol{x}_{\delta \widehat{\boldsymbol{z}}_{k-1}} = \boldsymbol{y}_{k-1} \cdot h'_\alpha \delta \widehat{\boldsymbol{z}}_{k-1} \cdot \boldsymbol{y}_{k-1} \cdot \boldsymbol{z}_k + \boldsymbol{y}_k \cdot \delta \widehat{\boldsymbol{z}}_{k-1} \cdot h_\alpha\left[\boldsymbol{x}_{k-1}\right] \\ + \boldsymbol{y}_k \cdot \boldsymbol{z}_{k-1} \cdot \boldsymbol{y}_{k-1} \cdot h'_\alpha \delta \widehat{\boldsymbol{z}}_{k-1}\,,\end{aligned} \tag{5.5}$$

[1] In this case, κ is double exponential with decreasing ζ.

and

$$\boldsymbol{x}_{\delta\widehat{\boldsymbol{y}}_{k-1}} = h_\alpha\left[\boldsymbol{x}_{k-1}\right]\cdot\delta\widehat{\boldsymbol{y}}_{k-1}\cdot\boldsymbol{z}_k + h'_\alpha\delta\widehat{\boldsymbol{y}}_{k-1}\cdot\boldsymbol{z}_{k-1}\cdot\boldsymbol{y}_{k-1}\cdot\boldsymbol{z}_k + \boldsymbol{y}_k\cdot\boldsymbol{z}_{k-1}\cdot h'_\alpha\delta\widehat{\boldsymbol{y}}_{k-1}\cdot\boldsymbol{z}_{k-1}\,. \quad (5.6)$$

Closer to the fixed point orbit, $\boldsymbol{y}_k\cdot\boldsymbol{z}_{k-1}\to\boldsymbol{I}$, $\boldsymbol{y}_{k-1}\cdot\boldsymbol{z}_k\to\boldsymbol{I}$, $h_\alpha\left[\boldsymbol{x}_k\right]\to\boldsymbol{I}$ and $h'_\alpha\to-\frac{1}{2}$ [67]. Then,

$$\boldsymbol{x}_{\delta\widehat{\boldsymbol{y}}_{k-1}}\to\delta\widehat{\boldsymbol{y}}_{k-1}\cdot\left(\boldsymbol{z}_k-\boldsymbol{z}_{k-1}\right) \quad (5.7)$$

and

$$\boldsymbol{x}_{\delta\widehat{\boldsymbol{z}}_{k-1}}\to\left(\boldsymbol{y}_k-\boldsymbol{y}_{k-1}\right)\cdot\delta\widehat{\boldsymbol{z}}_{k-1}. \quad (5.8)$$

Likewise, in the single channel instance:

$$\boldsymbol{x}_{\widehat{\boldsymbol{z}}_{k-1}}\to\left(\boldsymbol{z}_k-\boldsymbol{z}_{k-1}\right)^T\cdot\boldsymbol{s}\cdot\delta\widehat{\boldsymbol{z}}_{k-1}+\delta\widehat{\boldsymbol{z}}_{k-1}^T\cdot\boldsymbol{s}\cdot\left(\boldsymbol{z}_k-\boldsymbol{z}_{k-1}\right)\,. \quad (5.9)$$

About the fixed point then, error flow in the $\boldsymbol{y}$ and the $\boldsymbol{z}$ channels is tightly quenched, corresponding to $\boldsymbol{x}_{\delta\widehat{\boldsymbol{x}}_{k-1}}\to\boldsymbol{I}$ and identity iteration [67].

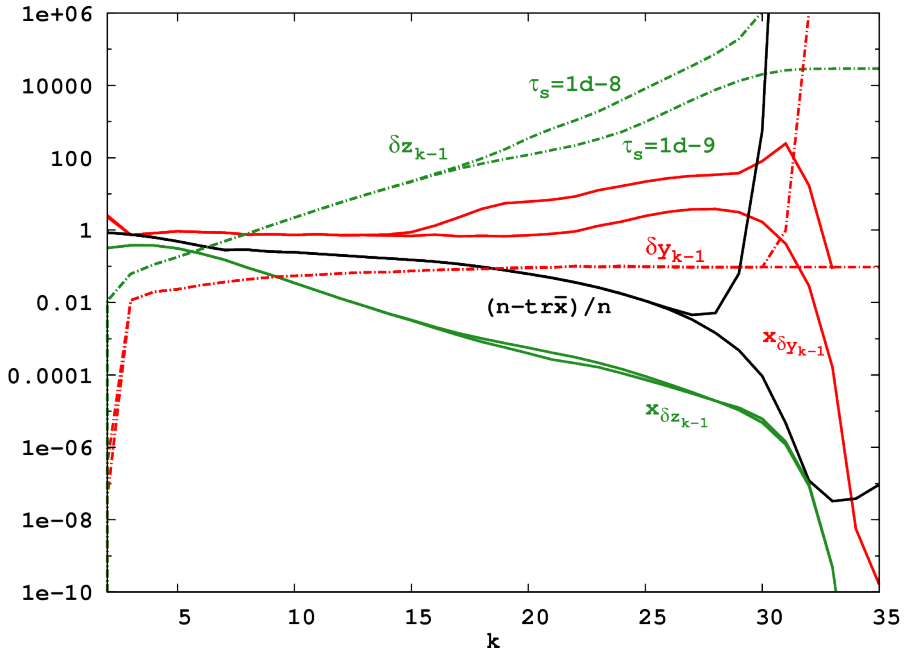

FIG. 5.1. *Trace error and $\|\cdot\|$ of derivatives and displacements for the unscaled dual iteration. Derivatives are full lines, whilst displacements for $\tau_s=\{10^{-8},10^{-9}\}$ are dashed lines. The trace error is a full black line.*

5.3. Displacements. At each step, the accumulation of previous errors in addition to the `SpAMM` occlusion error move the approximate iteration away from the unperturbed reference, here the `double-precision` iteration of arrays with `MATMUL`.

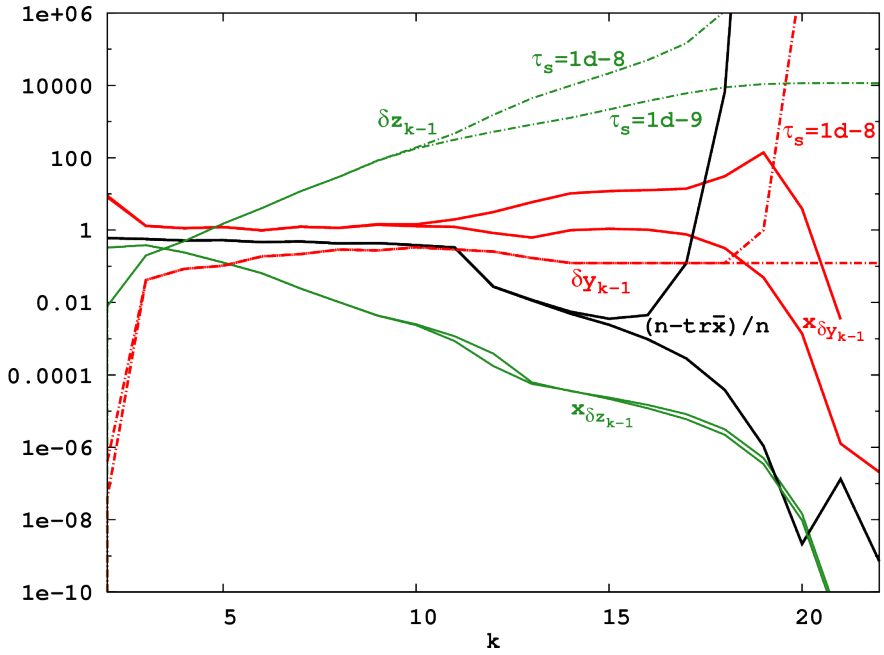

FIG. 5.2. *Trace error and $\|\cdot\|$ of derivatives and displacements for the scaled dual iteration. Derivatives are full lines, whilst displacements for $\tau_s = \{10^{-8}, 10^{-9}\}$ are dashed lines. The trace error is a full black line.*

Including the `SpAMM` error in the $\widetilde{\boldsymbol{z}}_{k-1}$ update we have:

$$\widetilde{\boldsymbol{z}}_{k-1} = \widetilde{\boldsymbol{z}}_{k-2} \otimes_\tau h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}] = \Delta_\tau^{\widetilde{\boldsymbol{z}}_{k-2}\cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]} + \widetilde{\boldsymbol{z}}_{k-2} \cdot h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] . \tag{5.10}$$

Then, with $h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] = h_\alpha\left[\boldsymbol{x}_{k-2}\right] + h'_\alpha \delta\boldsymbol{x}_{k-2}$, and taking $\boldsymbol{z}_{k-1}$ from both sides,

$$\delta\boldsymbol{z}_{k-1} = \Delta_\tau^{\widetilde{\boldsymbol{z}}_{k-2}\cdot h_\alpha[\widetilde{\boldsymbol{x}}_{k-2}]} + \delta\boldsymbol{z}_{k-2} \cdot h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right] + \boldsymbol{z}_{k-2} \cdot h'_\alpha \delta\boldsymbol{x}_{k-2} , \tag{5.11}$$

which is bounded by

$$\begin{aligned}\delta z_{k-1} < \|\boldsymbol{z}_{k-2}\| \left(\tau\, n^2 \left\|h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]\right\| + h'_\alpha \delta y_{k-2} \|z_{k-2}\|\right) \\ + \delta z_{k-2} \left(\left\|h_\alpha\left[\widetilde{\boldsymbol{x}}_{k-2}\right]\right\| + \|y_{k-2}\|\right) .\end{aligned} \tag{5.12}$$

In Eq. (5.12), the term $h'_\alpha \delta y_{k-2} \|\boldsymbol{z}_{k-2}\|^2$ is volatile, tending towards $\delta y_{k-2}\, \kappa(\boldsymbol{s})/2$. Because of this sensitivity, and because the $\boldsymbol{y}$ product channel maintains fidelity of the starting eigen-basis, we single out this "sensitive" product for a higher level of precision; $\tau_s \ll \tau$.

In the single instance, the $\boldsymbol{y}$ channel is implicit in the first product involving $\boldsymbol{s}$, which can be from the left or the right. In this work, the most accurate product in the single instance is rightmost.

5.4. Most approximate yet still stable. The potential to compute fast and effective preconditioners with `SpAMM` is determined by the most approximate yet still stable (MAYSS) iteration, a challenge for increasingly ill-conditioned problems. Illustrative experiments were carried out on the $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube examples

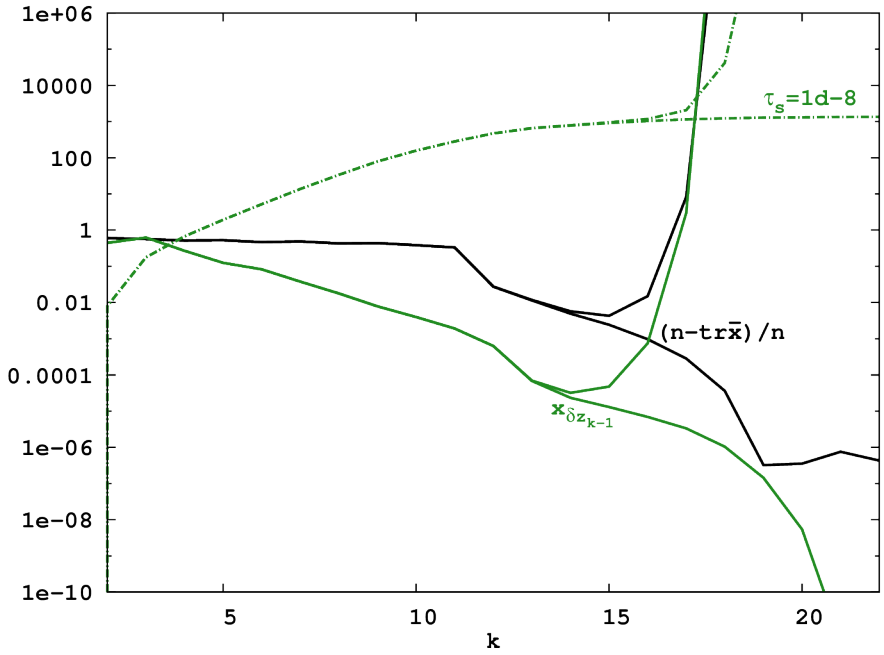

FIG. 5.3. *Trace error and $\|\cdot\|$ of derivatives and displacements for the scaled single iteration. Derivatives are full lines, whilst displacements for $\tau_s = \{10^{-7}, 10^{-8}\}$ are dashed lines. The trace error is a full black line.*

described in Section 4.3. We picked $\tau = .001$ and $N_b = 32$. Then, we looked at stability with respect to the tighter τ_s threshold: In Fig. 5.1, unscaled results for the dual instance are shown. In Fig. 5.2, scaled results for the dual instance are given, and in Fig. 5.3 we show results for the scaled single instance.

In the dual instances, Figs. 5.1 & 5.2, the bifurcating orientational components of the error manage to avoid the numerical catastrophe, with $\boldsymbol{x}_{\widehat{\boldsymbol{z}}_{k-1}}$ in solid green converging strongly, and $\boldsymbol{x}_{\widehat{\boldsymbol{y}}_{k-1}}$ in solid red, with an above unity drift driving divergence of the displacements (dashed lines). On the other hand, bifurcation in the single instance ($\tau_s = 10^{-7}$) finds the orientational component of the error, $\boldsymbol{x}_{\delta\widehat{\boldsymbol{z}}_{k-1}}$, diverging well ahead of the displacement δz_{k-1}.

In these problems, values of τ near the MAYSS bifurcation do not lead to a reduced complexity; instead, near total fill of the product is observed towards convergence, even for the largest 128× unit cell nano-tube. Also, scaling as reported in Section 4 reclaims about $\sim 2/3$ of the available 2× acceleration possible at this level off ill-conditioning [33], but dramatically enhances fill-in of the metric tensor, via the multiplicative effect of h'_α in Eq. (5.12). In addition to scaling, the single instance also results in a much larger volumetric fill-in, involving extended, delocalized error flows in the orbit.

Our interpretation of these results is that despite a similar overall convergence behavior and error control, the tensor volumes accessed by the two instances is very different, due to the magnitude of norms entering the `SpAMM` kernel; in the dual instance $y_k^{\text{dual}} = h[\boldsymbol{x_{k-1}}]\otimes_\tau \boldsymbol{y}_{k-1}$ is well behaved , while $y_k^{\text{single}} = \boldsymbol{z}_k^T \otimes_\tau \boldsymbol{s}$ encumbers large norms associated with the broad spectral resolution, leading to extended delocalization of the metric tensor.

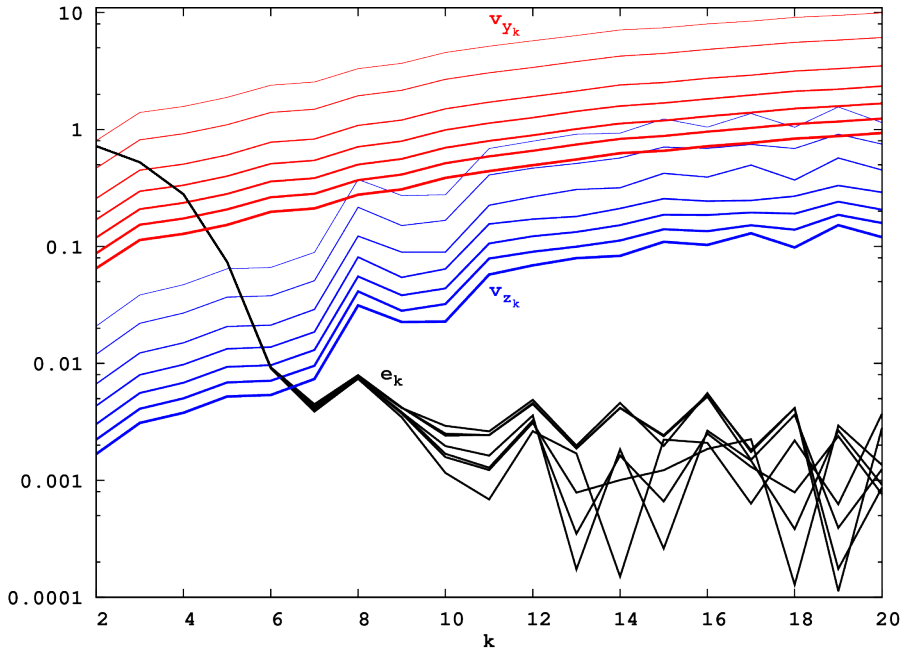

FIG. 5.4. *Culled volumes in the thin slice, single instance approximation of* $\boldsymbol{s}_{\tau_0\mu_0}^{-1/2}$ *for the* $\kappa(\boldsymbol{s}) = 10^{10}$ *nanotube sequence (line width increasing with system size). With* $\mu_0 = .1$ *it was only possible to achieve stability down to* $\tau_0 = 10^{-2}$ *&* $\tau_s = 10^{-4}$*. Shown are* $\mathrm{v}_{\widetilde{\boldsymbol{z}}_k} \equiv \left(\mathrm{vol}_{\widetilde{\boldsymbol{z}}_{k-1}\otimes_\tau \mathrm{h}[\widetilde{\boldsymbol{x}}_{k-1}]}\right) \times 100\%/\mathrm{n}^3$ *(blue) and* $\mathrm{v}_{\widetilde{\boldsymbol{y}}_k} \equiv \left(\mathrm{vol}_{\boldsymbol{s}\otimes_{\tau_s}\widetilde{\boldsymbol{z}}_k}\right) \times 100\%/n^3$ *(red). Also shown is the trace error,* $\mathrm{t}_k = (n - \mathrm{tr}\,\widetilde{\boldsymbol{x}}_k)/n$ *(black).*

6. Regularization. Even for the most approximate yet still stable approximations (MAYSS), our nanotube calculations lead to delocalized products that are not tightly bound by Eq. (2.3), even for very large 128× unit cell systems. And while similarly ill-conditioned problems may achieve substantial compression with just the MAYSS approximation, as shown later in Fig. 7.3, the `SpAMM` approximation cannot generally yield a fast method in cases of severe ill-conditioning.

A systematic way to reduce these effects is through Tikhonov regularization [92, 106]. Regularization involves a small level shift of the eigenvalues, $\boldsymbol{s}_\mu \leftarrow \boldsymbol{s} + \mu\boldsymbol{I}$, altering the condition number of the shifted matrix to $\kappa(\boldsymbol{s}_\mu) = \frac{\sqrt{s_{n-1}^2+\mu^2}}{\sqrt{s_0^2+\mu^2}}$ [106].

Achieving substantial acceleration with severe ill-conditioning may require a large level shift however, producing inverse factors of little practical use. One approach to recover a more accurate inverse factor is Riley's method based on Taylor's series [106];

$$\boldsymbol{s}^{-1/2} = \boldsymbol{s}_\mu^{-1/2} \cdot \left(\boldsymbol{I} + \frac{\mu}{2}\boldsymbol{s}_\mu^{-1} + \frac{3\mu^2}{8}\boldsymbol{s}_\mu^{-2} + \dots\right) . \tag{6.1}$$

For severely ill-conditioned problems and large level shifts, this expansion may converge very slowly. Also, adding powers of the full inverse may not be computationally effective.

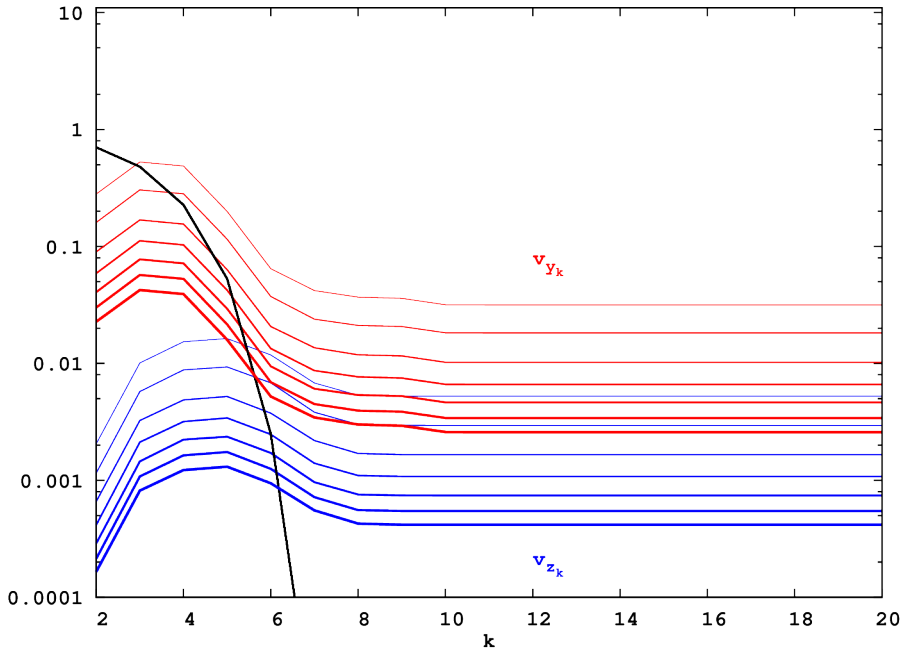

FIG. 5.5. *Culled volumes in the thin slice, dual instance approximation of* $\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}$ *for the* $\kappa(\boldsymbol{s}) = 10^{10}$ *nanotube sequence (line width increasing with system size). The slice is* $\mu_0 = .1, \tau_0 = .1$ *&* $\tau_s = .001$ *thin. Shown are* $\mathrm{v}_{\widetilde{\boldsymbol{y}}_k} = \left(\mathrm{vol}_{h[\widetilde{\boldsymbol{x}}_{k-1}]\otimes_{\tau_s}\widetilde{\boldsymbol{y}}_k}\right) \times 100\%/n^3$ *(blue) and* $\mathrm{v}_{\widetilde{\boldsymbol{z}}_k} = \left(\mathrm{vol}_{\widetilde{\boldsymbol{z}}_{k-1}\otimes_\tau h[\widetilde{\boldsymbol{x}}_{k-1}]}\right) \times 100\%/n^3$ *(red). Also shown is the trace error,* $\mathrm{t}_k = (n - \mathrm{tr}\,\widetilde{\boldsymbol{x}}_k)/n$ *(black), which rapidly approaches* 10^{-11} *(not shown).*

6.1. Product representation. We introduce an alternative representation of the regularized inverse factor;

$$\boldsymbol{s}^{-1/2} \equiv \bigotimes_{\substack{\tau=\tau_0\\ \mu=\mu_0}} |\,\tau\,\mu\,;\boldsymbol{s}^{-1/2}\rangle\,, \tag{6.2}$$

which is a telescoping product of preconditioned "slices" starting with a most-approximate-yet-still-effective-by-one-order (MAYEBOO) preconditioner, $\boldsymbol{s}^{-1/2}_{\tau_0\mu_0} \equiv |\,\tau_0\,\mu_0\,;\boldsymbol{s}^{-1/2}\rangle$ Braket notation marks the potential for asymmetries in the intermediate representation. This sandwich of generic, thinly sliced `SpAMM` products allows to construct a nested scoping on precision via τ, and in the effective condition number controlled by μ.

6.2. Effective by one order. We look again at the $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube series described in Section 4.3, this time with extreme regularization, $\mu_0 = .1$, and at a finer granularity, $N_b = 8$. Culled $\boldsymbol{y}_k$ and $\boldsymbol{z}_k$ volumes (as percentage of the total work) for $36 - 128\times$ the (3,3) unit cell are shown for the MAYEBOO approximation in Fig. 5.4 for the single instance, and in Fig. 5.5 for the dual instance.

The behavior of these implementations is very different; in the single instance, a stable iteration could not be found at precision $\tau_0 = .1$. Stability could only be found at .01, and that with a poorly contained trace error and cull-volumes that continue to inflate past convergence, with a conspicuous $\sqrt{k}$-like dependence. This behavior results

from the broad resolution of spectral powers $\widetilde{\boldsymbol{y}}_k \to \boldsymbol{s}^{-1/2}_{\tau_0\mu_0} \otimes_{\tau_0} \boldsymbol{s}_{\mu_0}$, with corresponding large metric fields that are poorly bound by Eq. (2.3).

On the other hand, dual iteration volumes collapse rapidly with fast trapping of the trace error, as $\widetilde{\boldsymbol{y}}_k \to \boldsymbol{I}_{\tau_0\mu_0} \otimes_{\tau_0} \boldsymbol{s}^{1/2}_{\tau_0\mu_0}$ and $\widetilde{\boldsymbol{z}}_k \to \boldsymbol{s}^{-1/2}_{\tau_0\mu_0} \otimes_{\tau_0} \boldsymbol{I}_{\tau_0\mu_0}$, and with Eq. (2.3) tightening to

$$\Delta^{\boldsymbol{I}_{\tau_0\mu_0} \cdot \boldsymbol{s}^{1/2}_{\tau_0\mu_0}} < \tau n \| \boldsymbol{s}^{1/2}_{\tau_0\mu_0} \| \tag{6.3}$$

and

$$\Delta^{\boldsymbol{s}^{-1/2}_{\tau_0\mu_0} \cdot \boldsymbol{I}_{\tau_0\mu_0}} < \tau n \| \boldsymbol{s}^{-1/2}_{\tau_0\mu_0} \| \,. \tag{6.4}$$

This contraction to the plane diagonal is compressive, leading to computational complexities that should approach quadtree copy in place.

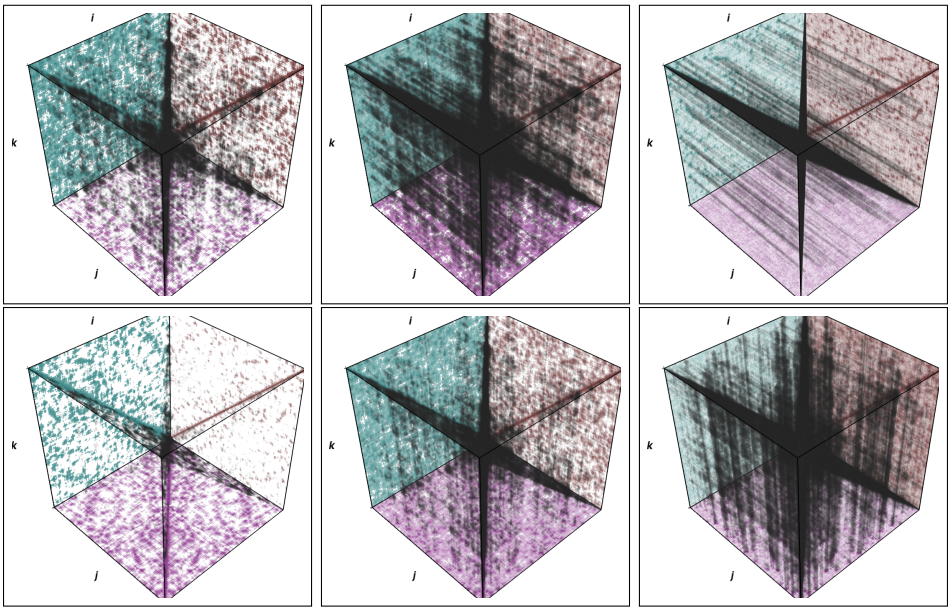

FIG. 6.1. *Product volumes in construction of the unregularized preconditioner* $|\tau_0 = .001; \boldsymbol{s}^{-1/2}\rangle$, *with dual instance square root iteration, and for the 6-311G** metric of 100 periodic water molecules. At top its* $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}] \otimes_{\tau_s} \boldsymbol{y}_{k-1}$ *for* $k = 0, 5, \&17$, *while on the bottom we have* $\boldsymbol{x}_k = \boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k$ *for* $k = 0, 5, \&17$. *Maroon is* $\boldsymbol{a}$, *purple is* $\boldsymbol{b}$, *green is* $\boldsymbol{c}$, *and black is the volume* $\mathrm{vol}_{a \otimes_\tau b}$ *in the product* $\boldsymbol{c} = \boldsymbol{a} \otimes_\tau \boldsymbol{b}$.

6.3. Iterative regularization. We now sketch an iterative approach to constructing the product representation, Eq. (6.2). In the dual instance, it is possible to obtain a first MAYEBOO approximation $\boldsymbol{s}^{-1/2}_{\tau_0=.1,\mu_0=.1}$, which improves the condition number by one order of magnitude, with a numerical resolution of approximately one digit. Then, a next level slice can be found, $\boldsymbol{s}^{-1/2}_{\tau_0\mu_1}$, based on the residual $\left(\boldsymbol{s}^{-1/2}_{\tau_0\mu_0}\right)^T \otimes_{\tau_1} (\boldsymbol{s} + \mu_1 \boldsymbol{I}) \otimes_{\tau_1} \boldsymbol{s}^{-1/2}_{\tau_0\mu_0}$, with *e.g.* $\mu_1 = .01$ and $\tau_1 = .01$. The product $\boldsymbol{s}^{-1/2}_{\tau_0\mu_1} \otimes_{\tau_1} \boldsymbol{s}^{-1/2}_{\tau_0\mu_0}$ then improves the condition number by two orders of magnitude, still with a numerical resolution of one digit. Reflected in the preceding notation, it appears necessary to compute the residual at a higher level of precision, *e.g.* using $\otimes_{\tau_1}$ instead of $\otimes_{\tau_0}$ and with $\tau_1 > \tau_0$.

In this way, it may be possible to obtain product representation of the inverse square root at a SpAMM resolution that is potentially far more permissive than otherwise possible,

$$\boldsymbol{s}_{\tau_0}^{-1/2} = \boldsymbol{s}_{\tau_0\mu_n}^{-1/2} \otimes_{\tau_1} \boldsymbol{s}_{\tau_0\mu_{n-1}}^{-1/2} \otimes_{\tau_1} \ldots \boldsymbol{s}_{\tau_0\mu_0}^{-1/2}, \tag{6.5}$$

assuming $.1 \geq \mu_0 > \mu_1 > \ldots$ Likewise, it may also be possible to obtain the full inverse factor with increasing numerical resolution as

$$\boldsymbol{s}^{-1/2} = \boldsymbol{s}_{\tau_m}^{-1/2} \otimes_{\tau_{m+1}} \boldsymbol{s}_{\tau_{m-1}}^{-1/2} \otimes_{\tau_m} \ldots \boldsymbol{s}_{\tau_0}^{-1/2}, \tag{6.6}$$

and $.1 \geq \tau_0 > \tau_1 > \ldots$

Also, with each step a well conditioned generic slice, it may be possible to find a more effective logistic map optimized for a vanilla distribution of eigenvalues. Finally, relative to the regularization and precision scoping sketched here, alternative products are possible that may be far more efficient. We hope to pursue these efforts in future work.

7. Locality.

7.1. Spatial and metric locality. Astrophysical n-body algorithms employ range queries over spatial databases to hierarchically discover and compute approximations that commit only small errors. Often, these spatial databases are ordered with a space filling curve (SFC) [125, 1, 101, 124, 126, 84, 102, 2, 19, 7], which maps points that are close in space to an index where they are also close. Spatial locality of this type empowers the SpAMM approximation through Block-By-Magnitude orderings of the sub-space metric.

This metric locality is compressive, but diminished by dimensionality. In Figure 6.1, we show $\otimes_\tau$ volumes for square root iteration, corresponding to the Gramian matrix of a small, periodic water box with the large 6-311G** basis (Section 4.3). In this 3-d periodic case, diminishing Cartesian separations lead to long-skinny pillae and related delocalizations not observed in lower dimensional problems at this modest $\kappa(\boldsymbol{s}) \sim 10^5$ level of ill-conditioning. These delocalizations correspond to weakness in Eq. (2.3), and to tighter values of τ_s, required in the MAYSS approximation. As n becomes large, Cartesian separation will eventually thin these delocalizations leading to complexity reduction due only to metric decay.

7.2. Algebraic locality. In addition to compression through orderings that maximize these block-by-magnitude effects, we demonstrate a new kind of locality in Figs. 7.1 and 7.2, which is, so far, uniquely exploited by the n-body approach to square root iteration. This locality increases compressively towards convergence, as contractive identity iterations develop. We call this compression ***lensing***, involving collapse of the culled volume about plane diagonals of the identity. Lensing corresponds to strengthening Eq.(2.3), viz Eqs. (6.3)-(6.4), and to strongly contracting directional derivatives, viz Eqs. (5.7)-(5.8). This is an important, mitigating effect for SpAMM computations in the $\boldsymbol{y}$ channel, encumbered by the parameter $\tau_s \sim .01 \times \tau$.

Graph reorderings that minimize the distance of matrix elements from the diagonal also lead to matrix locality (aforementioned). In Fig. 7.3 we show convergence of an unregularized (MAYSS) preconditioner for this type of ordering and the bcsstk14 [123] structural matrix of the Omni Coliseum in Atlanta, with $\kappa(\boldsymbol{s}) = 10^{10}$. These results show remarkable gossamer sheeting and flattening along plane diagonals, in Fig. 7.3, at top for development of $\boldsymbol{y}_k$, as well as hollow accumulation of $\mathrm{vol}_{\boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k}$

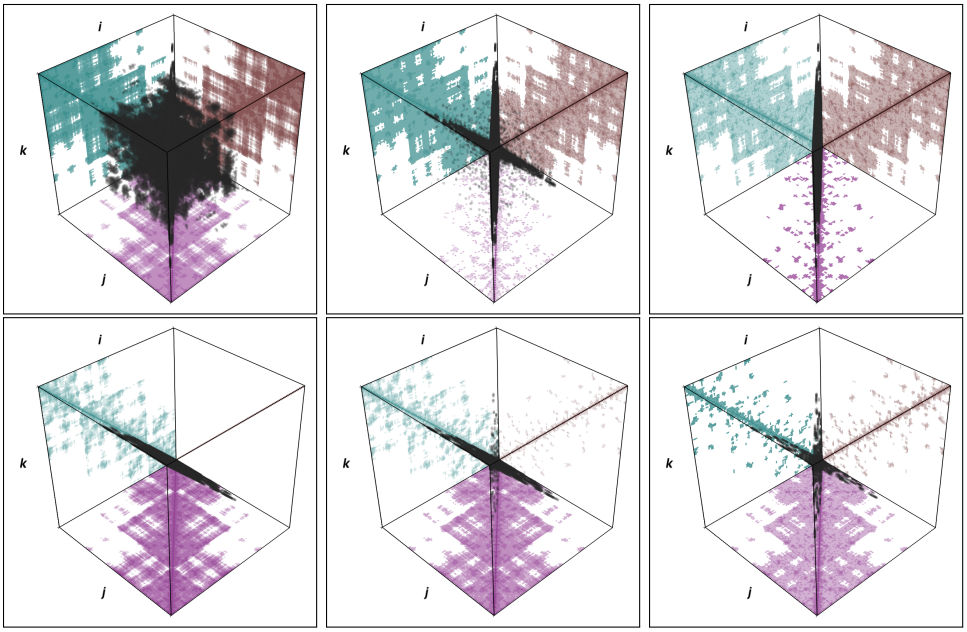

FIG. 7.1. *Product volumes in construction of the MAYEBOO preconditioner* $|\tau_0 = .1, \mu_0 = .1\,; \boldsymbol{s}^{-1/2}\rangle$*, with dual instance square root iteration, for* $8\times$ $\kappa(s) = 10^{11}$ *nano-tube.* $\boldsymbol{y}_k$ *appears wider than* $\boldsymbol{z}_k$ *because it is computed at a higher precision,* $\tau_s = .001$*, and because the first multiply involves* $\boldsymbol{s}^2$*. At top its* $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s}\boldsymbol{y}_{k-1}$ *for* $k = 0, 4, \&16$*, while on the bottom we have* $\boldsymbol{x}_k = \boldsymbol{y}_k\otimes_\tau \boldsymbol{z}_k$ *for* $k = 0, 2, \&16$*. Maroon is* $\boldsymbol{a}$*, purple is* $\boldsymbol{b}$*, green is* $\boldsymbol{c}$*, and black is the volume* $\mathrm{vol}_{a\otimes_\tau b}$ *in the product* $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau \boldsymbol{b}$*.*

at bottom. Interestingly, this example demonstrates well lensed volumes towards convergence, whilst the equally ill-conditioned and lower dimensional $\kappa(\boldsymbol{s}) = 10^{10}$ nanotube demands a much tighter value of τ_s (10^{-4} vs. 10^{-9}) and retains dense volumes through $128\times$ the unit cell.

7.3. Complexity reduction. Finally, we show complexity reduction at convergence of the MAYEBOO approximation relative to the MAYSS approximation, in Fig. 7.4 for periodic water boxes, and in Fig. 7.5 for the ill-conditioned nano-tube. The two-orders difference between $\boldsymbol{y}_k$ and $\boldsymbol{z}_k$ volumes corresponds precisely to $\tau_s \sim \tau \times .01$, with $\boldsymbol{x}_k$ in between. Except for the slower trend in Fig. (7.4)'s $\boldsymbol{x}_k$ volume, we see the potential for continued strong acceleration with increasing system size.

8. Conclusions. In this work, we developed the `SpAMM` n-body solver for square root iteration, along with some algebra for the operator $\otimes_\tau$, and showed how we could exploit different types of locality in the sub-space metric of the product-tensor. Our main contributions include a modified Cauchy-Schwarz criterion for the `SpAMM` occlusion-cull, Eq. (2.2), and proof that the corresponding relative error in the product is bound by Eq. (2.3). We showed how block-by-magnitude orderings and locality of the sub-space metric leads to reduced complexity of the `SpAMM` kernel, involving low-dimensional sub-structures that bound the relative error, distributed along plane-diagonals and along their their intersection at the cube-diagonal. Perhaps most significantly, we demonstrated a new kind of compressive locality, lensing, that develops in the $\otimes_\tau$ volume on contractive identity iteration, together with tightening the `SpAMM` bound, viz Eqs. (6.3)-(6.4).

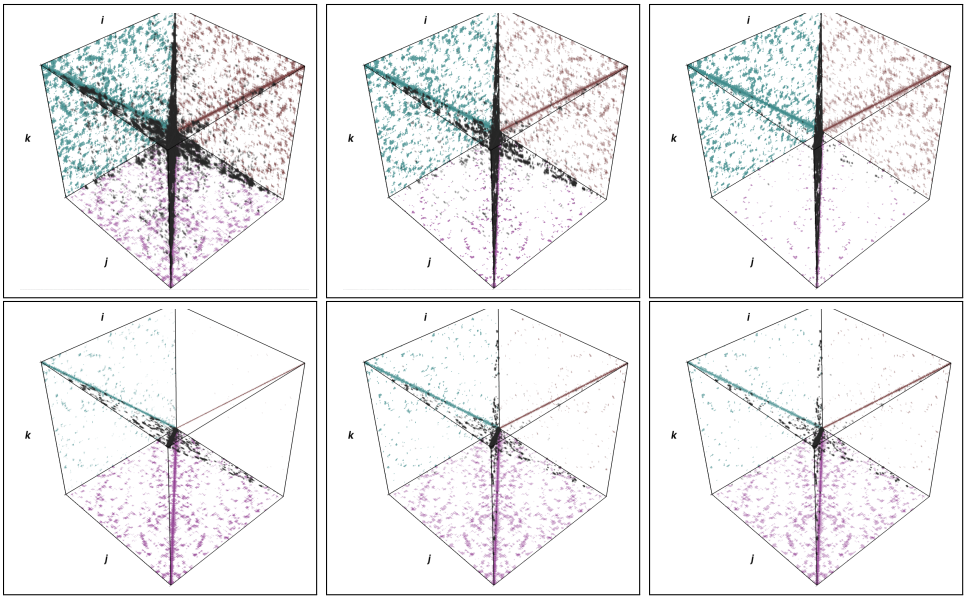

FIG. 7.2. *Product volumes in construction of the MAYEBOO preconditioner* $|\tau_0 = .1, \mu_0 = .1\,; \boldsymbol{s}^{-1/2}\rangle$*, with dual instance square root iteration, for 6-311G** box of 100 periodic water molecules. At top its* $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s}\boldsymbol{y}_{k-1}$ *for* $k = 0, 4, \&15$*, while on the bottom we have* $\boldsymbol{x}_k = \boldsymbol{y}_k \otimes_\tau \boldsymbol{z}_k$ *for* $k = 0, 4, \&15$*. Maroon is* $\boldsymbol{a}$*, purple is* $\boldsymbol{b}$*, green is* $\boldsymbol{c}$*, and black is the volume* $\mathrm{vol}_{a\otimes_\tau b}$ *in the product* $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau \boldsymbol{b}$*.*

Additional contributions include development and implementation of a first order Fréchet analyses for the single and dual instances of the NS square root iteration, with focus on separating directional effects that are mostly controlled by the unperturbed reference algebra, from the magnitude of `SpAMM` occlusion errors and their accumulation. We found that numerical sensitivity develops primarily in the $\boldsymbol{z}$ channel, according to Eq. (5.12), due to amplification of δy by terms approaching condition of the full inverse; we then looked at sensitivity to this error, bifurcations, controled by τ_s (Figs. 5.1-5.3), concluding that a most approximate, naive application of `SpAMM` to the ill-conditioned problem is generally insufficient to achieve a fast solution.

Finally, we introduced scoping on both precision and regularization in product representation of the inverse factor, and demonstrated the potential for orders of magnitude compression in the dual instance, Figs. 7.4-7.5, with the most extreme, "by-one-order" slice of the nested factor, providing a foothold for this expansion at $\tau_0 = .1$. A next step is to demonstrate full bootstrapping of the inverse factor with reduced complexity, *i.e.* via a compact, nested product of well lensed terms, a work in progress.

REFERENCES

[1] S. KAMAL ABDALI AND DAVID S. WISE, *Experiments with quadtree representation of matrices*, in Symbolic and Algebraic Computation, P. Gianni, ed., vol. 358 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 1989, pp. 96–108.

[2] MICHAEL D. ADAMS AND DAVID S. WISE, *Seven at one stroke: results from a cache-oblivious paradigm for scalable matrix algorithms*, in Proceedings of the 2006 workshop on Memory system performance and correctness, MSPC '06, New York, NY, USA, 2006, ACM, pp. 41–50.

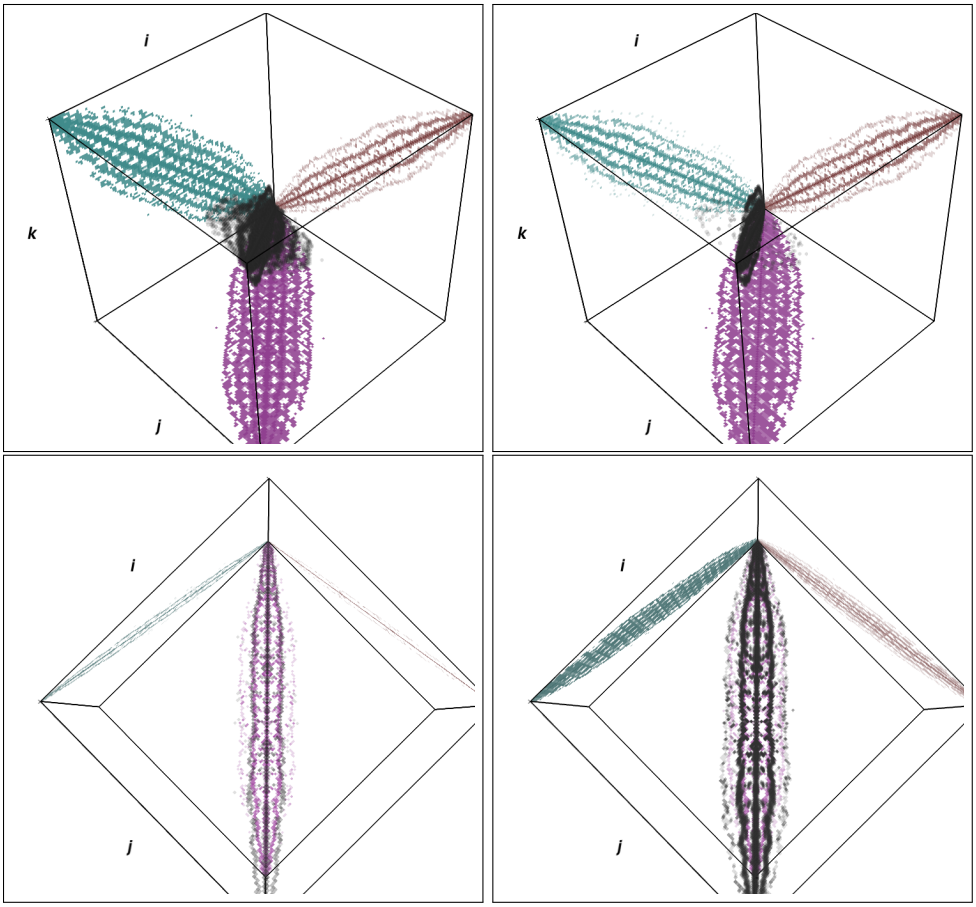

FIG. 7.3. *Product volumes in construction of the unregularized preconditioner* $|\tau_0 = .001, \mu_0 = .0\,; \boldsymbol{s}^{-1/2}\rangle$*, with the dual instance of square root iteration and for the* `bcsstk14` *structural matrix. At top its* $\boldsymbol{y}_k = h_\alpha[\boldsymbol{x}_{k-1}]\otimes_{\tau_s}\boldsymbol{y}_{k-1}$ *for* $k = 0$ *&* 37*, while on the bottom we have* $\boldsymbol{x}_k = \boldsymbol{y}_k\otimes_\tau \boldsymbol{z}_k$ *for* $k = 0$ *&* 37*. Maroon is* $\boldsymbol{a}$*, purple is* $\boldsymbol{b}$*, green is* $\boldsymbol{c}$*, and black is the volume* $\mathrm{vol}_{a\otimes_\tau b}$ *in the product* $\boldsymbol{c} = \boldsymbol{a}\otimes_\tau \boldsymbol{b}$*.*

[3] K AKBUDAK, ENVER KAYAASLAN, AND CEVDET AYKANAT, *Technical Report on Hypergraph-Partitioning-Based Models and Methods for Exploiting Cache Locality in Sparse-Matrix Vector Multiplication Kadir Akbudak , Enver Kayaaslan and Cevdet Aykanat February 2012*, (2012).

[4] R. R. AMOSSEN AND R. PAGH, *Faster join-projects and sparse matrix multiplications*, in Proceedings of the 12th International Conference on Database Theory, ACM, 2009, pp. 121–126.

[5] YUVAL FILMUS ANDRIS AMBAINIS AND FRANOIS LE GALL, *Fast matrix multiplication: Limitations of the coppersmith-winograd method*, in Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC '15, New York, NY, USA, 2015, ACM, pp. 585–593.

[6] LUC ANSELIN, *Spatial Externalities, Spatial Multipliers, And Spatial Econometrics*, Int. Reg. Sci. Rev., 26 (2003), pp. 153–166.

[7] MICHAEL BADER, *Space-Filling Curves: An Introduction with Applications in Scientific Computing*, vol. 9 of Texts in Computational Science and Engineering, Springer Berlin Heidelberg, 2013.

[8] ZHAOJUN BAI AND JAMES DEMMEL, *Using the matrix sign function to compute invariant subspaces*, SIAM J. Matrix Anal. Appl, 19 (1998), pp. 205–225.

[9] GREY BALLARD, JAMES DEMMEL, OLGA HOLTZ, AND ODED SCHWARTZ, *Communication costs*

of Strassen's matrix multiplication, Communications of the ACM, 57 (2014), pp. 107–114.
[10] Mikhail Belkin and Partha Niyogi, *Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering*, Science (80-.)., 1 (2002), pp. 585–591.
[11] ———, *Laplacian Eigenmaps for Dimensionality Reduction and Data Representation*, Neural Comput., 15 (2003), pp. 1373–1396.
[12] ———, *Towards a theoretical foundation for Laplacian-based manifold methods*, J. Comput. Syst. Sci., 74 (2008), pp. 1289–1308.
[13] Michele Benzi, *Preconditioning Techniques for Large Linear Systems: A Survey*, J. Comput. Phys., 182 (2002), pp. 418–477.
[14] Michele Benzi, Paola Boito, and N Razouk, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Rev., 55 (2013), pp. 3–64.
[15] Michele Benzi and Gene H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–438.
[16] Michele Benzi and Nader Razouk, *Decay Bounds and $\mathcal{O}(n)$ Algorithms for Approximating Functions of Sparse Matrices*, Electron. T. Numer. Ana., 28 (2007), p. 16.
[17] Michele Benzi and M Tuma, *A comparative study of sparse approximate inverse preconditioners*, Appl. Numer. Math., 30 (1999), pp. 305–340.
[18] Åke Björck, *The calculation of linear least squares problems*, Acta Numerica, 13 (2004), pp. 1–53.
[19] Nicolas Bock and Matt Challacombe, *An Optimized Sparse Approximate Matrix Multiply for Matrices with Decay*, SIAM Journal on Scientific Computing, 35 (2013), pp. C72–C98.
[20] Nicolas Bock, Matt Challacombe, and Laxmikant V. Kalé, *Solvers for $\mathcal{O}(N)$ Electronic Structure in the Strong Scaling Limit*, (2014).
[21] D. R. Bowler and T. Miyazakia andM. J. Gillan, *Parallel sparse matrix multiplication for linear scaling electronic structure calculations*, Comp. Phys. Comm., 137 (2000), pp. 255–273.
[22] D R Bowler and T Miyazaki, *$\{O\}(N)$ methods in electronic structure calculations*, Reports Prog. Phys., 75 (2012), p. 36503.
[23] V Brzdov and D R Bowler, *Automatic data distribution and load balancing with space-filling curves: implementation in conquest*, Journal of Physics: Condensed Matter, 20 (2008), p. 275223.
[24] Aydin Buluc, Henning Meyerhenke, Ilya Safro, Peter Sanders, and Christian Schulz, *Recent advances in graph partitioning*, arXiv preprint arXiv:1311.3144, (2013).
[25] Ralph Byers, Chunyang He, and Volker Mehrmann, *The matrix sign function method and the computation of invariant subspaces*, SIAM Journal on Matrix Analysis and Applications, 18 (1997), pp. 615–632.
[26] Claudio Canuto, Valeria Simoncini, and Marco Verani, *On the decay of the inverse of matrices that are sum of kronecker products*, Linear Algebra and its Applications, 452 (2014), pp. 21–39.
[27] Matt Challacombe, *A simplified density matrix minimization for linear scaling self-consistent field theory*, The Journal of chemical physics, 110 (1999), pp. 2332–2342.
[28] ———, *A general parallel sparse-blocked matrix multiply for linear scaling SCF theory*, Comput. Phys. Commun., 128 (2000), pp. 93–107.
[29] Matt Challacombe, *A general parallel sparse-blocked matrix multiply for linear scalingSCF theory.*, Comp. Phys. Comm., 128 (2000), p. 93.
[30] Matt Challacombe, *Linear Scaling Solution of the Time-Dependent Self-Consistent-Field Equations*, Computation, 2 (2014), pp. 1–11.
[31] Matt Challacombe, *Fast field transforms for free mesh interpolation.* 2016.
[32] Matt Challacombe and Nicolas Bock, *Fast Multiplication of Matrices with Decay*, arXiv:cs.DS, 1011.3534 (2010).
[33] Jie Chen and Edmond Chow, *A Newton-Schulz Variant for Improving the Initial Convergence in Matrix Sign Computation*, Preprint ANL/MCS-P5059-0114, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, 2014.
[34] Shimin Chen, Anastassia Ailamaki, Phillip B. Gibbons, and Todd C. Mowry, *Improving hash join performance through prefetching*, ACM Trans. Database Syst., 32 (2007).
[35] Eric Chisolm and Nicolas Bock, *Matrixtree*. `https://github.com/FreeON/spammpack/tree/master/src-Haskell`, 2015.
[36] Jiawei Chiu and Laurent Demanet, *Matrix probing and its conditioning*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 171–193.
[37] Henry Cohn, Robert Kleinberg, Balazs Szegedy, and Christopher Umans, *Group-theoretic algorithms for matrix multiplication*, in Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on, IEEE, 2005, pp. 379–388.

[38] Henry Cohn and Christopher Umans, *A group-theoretic approach to fast matrix multiplication*, in Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on, IEEE, 2003, pp. 438–449.

[39] Gianna M. Del Corso, Antonio Gullí, and Francesco Romani, *Fast PageRank Computation via a Sparse Linear System*, Internet Math., 2 (2005), pp. 251–273.

[40] Stephen Demko, William F Moss, and Philip W Smith, *Decay rates for inverses of band matrices*, Mathematics of computation, 43 (1984), pp. 491–499.

[41] James Demmel, Ioana Dumitriu, and Olga Holtz, *Fast linear algebra is stable*, Numerische Mathematik, 108 (2007), pp. 59–91.

[42] James W. Demmel, Ioana Dumitriu, Olga Holtz, and Robert Kleinberg, *Fast matrix multiplication is stable*, Numer. Math., 106 (2007), pp. 199–224.

[43] James Dinan, Sriram Krishnamoorthy, D Brian Larkins, Jarek Nieplocha, and Ponnuswamy Sadayappan, *Scioto: A framework for global-view task parallelism*, in Parallel Processing, 2008. ICPP'08. 37th International Conference on, IEEE, 2008, pp. 586–593.

[44] Petros Drineas, Ravi Kannan, and Michael W. Mahoney, *Fast Monte Carlo Algorithms for Matrices I: Approximating Matrix Multiplication*, SIAM Journal on Computing, 36 (2006), p. 132.

[45] Michael Driscoll, Evangelos Georganas, Penporn Koanantakool, Edgar Solomonik, and Katherine Yelick, *A Communication-Optimal N-Body Algorithm for Direct Interactions*, in Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on, IEEE, 2013, pp. 1075–1084.

[46] Erik Elmroth, Fred Gustavson, Isak Jonsson, and B Kå gström, *Recursive blocked algorithms and hybrid data structures for dense matrix library software*, SIAM Rev., 46 (2004), pp. 3–45.

[47] Massimo Fornasier and Karlheinz Gröchenig, *Intrinsic localization of frames*, Constructive Approximation, 22 (2005), pp. 395–415.

[48] Michael Frasca, Kamesh Madduri, and Padma Raghavan, *NUMA-aware graph mining techniques for performance and energy efficiency*, 2012 Int. Conf. High Perform. Comput. Networking, Storage Anal., (2012), pp. 1–11.

[49] J. T. Frey and D. J. Doren, *Tubegen 3.4*. `http://turin.nss.udel.edu/research/tubegenonline.html`, 2011.

[50] Floris Geerts, Sofie Haesevoets, and Bart Kuijpers, *A theory of spatio-temporal database queries*, in Database Programming Languages, Giorgio Ghelli and Gsta Grahne, eds., vol. 2397 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2002, pp. 198–212.

[51] Gene H. Golub and Charles F. Van Loan, *Matrix Computations (3rd Ed.)*, Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[52] Peter Gottschling, DS Wise, and A Joshi, *Generic support of algorithmic and structural recursion for scientific computing*, Int. J. Parallel, ..., (2009).

[53] Alexander Gray and Andrew Moore, *'n-body' problems in statistical learning*, in Advances in Neural Information Processing Systems 13, MIT Press, 2000, pp. 521–527.

[54] A. G. Gray, *Bringing Tractability to Generalized N-Body Problems in Statistical and Scientific Computation*, PhD thesis, 2003.

[55] Philipp Grohs, *Intrinsic localization of anisotropic frames*, Applied and Computational Harmonic Analysis, 35 (2013), pp. 264–283.

[56] Philipp Grohs and Stefano Vigogna, *Intrinsic localization of anisotropic frames ii:\ alpha-molecules*, Journal of Fourier Analysis and Applications, 21, pp. 182–205.

[57] Fred Gustavson, Isak Jonsson, Bo Kgstrm, and Per Ling, *Towards peak performance on hierarchical memory architectures - new recursive blocked data formats and blas's*, in Proceedings of the Ninth SIAM Conference on Parallel Processing for Scientific Computing, B. Hendrickson et. al., ed., PPSC99, SIAM, 1999.

[58] Fred G Gustavson, *Two fast algorithms for sparse matrices: Multiplication and permuted transposition*, ACM Transactions on Mathematical Software (TOMS), 4 (1978), pp. 250–269.

[59] Aleksandar Haber and Michel Verhaegen, *Sparse approximate inverses of gramians and impulse response matrices of large-scale interconnected systems*, CoRR, abs/1405.2580 (2014).

[60] Johanna Hardin, Stephan Ramon Garcia, and David Golan, *A method for generating realistic correlation matrices*, Ann. Appl. Stat., 7 (2013), pp. 1733–1762.

[61] Martin Head-Gordon, Michael Lee, Paul Maslen, Troy Van Voorhis, Steven Gwaltney, et al., *Tensors in electronic structure theory: basic concepts and applications to electron correlation models*.

[62] G Heber, R Biswas, and GR Gao, *Self-avoiding walks over two-dimensional adaptive unstructured grids*, 9 (1998).
[63] Christopher Heil, *History and evolution of the density theorem for gabor frames*, Journal of Fourier Analysis and Applications, 13 (2007), pp. 113–166.
[64] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-structure Theory*, Wiley, 2008.
[65] Nicholas J. Higham, *Accuracy and Stability of Numerical Algorithms*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd ed., 2002.
[66] ———, *Functions of Matrices*, Society for Industrial & Applied Mathematics,, 2008.
[67] Nicholas J. Higham, D. Steven Mackey, Niloufer Mackey, and Fracoise Tisseur, *Functions preserving matrix groups and iterations for the matrix square root*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 849–877.
[68] E. G. Hoel and H. Samet, *Data-Parallel Spatial Join Algorithms*, in Parallel Processing, 1994. ICPP 1994. International Conference on, vol. 3, aug. 1994, pp. 227 –234.
[69] Mark A Iwen and Craig V Spencer, *A note on compressed sensing and the complexity of matrix multiplication*, Information Processing Letters, 109 (2009), pp. 468–471.
[70] Edwin H. Jacox and Hanan Samet, *Iterative spatial join*, ACM Trans. Database Syst., 28 (2003), pp. 230–256.
[71] Branislav Jansík, Stinne Host, Poul Jorgensen, Jeppe Olsen, and Trygve Helgaker, *Linear-scaling symmetric square-root decomposition of the overlap matrix.*, J. Chem. Phys., 126 (2007), p. 124104.
[72] Kahan, *A Tutorial Overview of Vector and Matrix Norms*, (2013), pp. 1–79.
[73] Sepandar Kamvar, Taher Haveliwala, Christopher Manning, and Gene Golub, *Exploiting the block structure of the web for computing pagerank*, Stanford University Technical Report, (2003).
[74] Changkyu Kim, Tim Kaldewey, Victor W. Lee, Eric Sedlar, Anthony D. Nguyen, Nadathur Satish, Jatin Chhugani, Andrea Di Blas, and Pradeep Dubey, *Sort vs. Hash revisited: fast join implementation on modern multi-core CPUs*, Proc. VLDB Endow., 2 (2009), pp. 1378–1389.
[75] Ilya Krishtal, Thomas Strohmer, and Tim Wertz, *Localization of matrix factorizations*, Foundations of Computational Mathematics, 15 (2015), pp. 931–951.
[76] ———, *Localization of matrix factorizations*, Found. Comp. Math., 15 (2015), pp. 931–951.
[77] Gabriel Kuper, Leonid Libkin, and Jan Paredaens, eds., *Genericity in Spatial Databases*, 2000.
[78] Ilya Lashuk, Aparna Chandramowlishwaran, Harper Langston, Tuan-Anh Nguyen, Rahul Sampath, Aashay Shringarpure, Richard Vuduc, Lexing Ying, Denis Zorin, and George Biros, *A massively parallel adaptive fast-multipole method on heterogeneous architectures*, in SC '09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, New York, NY, USA, 2009, ACM, pp. 1–12.
[79] François Le Gall, *Powers of tensors and fast matrix multiplication*, in Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation, ISSAC '14, New York, NY, USA, 2014, ACM, pp. 296–303.
[80] Sylvain Lefebvre and Hugues Hoppe, *Perfect spatial hashing*, ACM Trans. Graph., 25 (2006), p. 579.
[81] Daniel Lemire, Owen Kaser, and Eduardo Gutarra, *Reordering rows for better compression*, ACM Trans. Database Syst., 37 (2012), pp. 1–29.
[82] Pei Li, Yuanzhe Cai, Hongyan Liu, Jun He, and Xiaoyong Du, *Exploiting the block structure of link graph for efficient similarity computation*, in Advances in Knowledge Discovery and Data Mining, Springer, 2009, pp. 389–400.
[83] Michael D. Lieberman, Jagan Sankaranarayanan, and Hanan Samet, *A Fast Similarity Join Algorithm Using Graphics Processing Units*, in Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Washington, DC, USA, 2008, IEEE Computer Society, pp. 1111–1120.
[84] K. Patrick Lorton and David S. Wise, *Analyzing block locality in morton-order and morton-hybrid matrices*, in Proceedings of the 2006 workshop on Memory performance: Dealing with Applications, systems and architectures, MEDEA '06, New York, NY, USA, 2006, ACM, pp. 5–12.
[85] P. O. Lowdin, Advances in Physics, 5 (1956), p. 1.
[86] Michael W Mahoney, *Implementing Randomized Matrix Algorithms in Parallel and Distributed Environments*, (2012), pp. 1–53.
[87] Qingyu Meng and Martin Berzins, *Scalable large-scale fluid–structure interaction solvers in the uintah framework via hybrid task-based parallelism algorithms*, Concurrency and

Computation: Practice and Experience, 26 (2014), pp. 1388–1407.
[88] Qingyu Meng, Justin Luitjens, and Martin Berzins, *Dynamic task scheduling for the uintah framework*, in Many-Task Computing on Grids and Supercomputers (MTAGS), 2010 IEEE Workshop on, IEEE, 2010, pp. 1–10.
[89] Seung-Jai Min, Costin Iancu, and Katherine Yelick, *Hierarchical work stealing on manycore clusters*, in 5th Conf. on Partitioned Global Address Space Prog. Models, 2011.
[90] Priti Mishra and Margaret H. Eich, *Join processing in relational databases*, ACM Comput. Surv., 24 (1992), pp. 63–113.
[91] Annavarapu Ramesh Naidu, *Centrality of Lowdin orthogonalizations*, May 2011.
[92] Arnold Neumaier, *Solving ill-conditioned and singular linear systems: A tutorial on regularization*, SIAM review, 40 (1998), pp. 636–666.
[93] Leonid Oliker, Xiaoye Li, Parry Husbands, and Rupak Biswas, *Effects of Ordering Strategies and Programming Paradigms on Sparse Matrix Computations*, SIAM Rev., 44 (2002), pp. 373–393.
[94] Rasmus Pagh, *Compressed matrix multiplication*, ACM Trans. Comput. Theory, 5 (2013), pp. 1–17.
[95] V Pan and R Schreiber, *An improved Newton iteration for the generalized inverse of a matrix, with applications*, SIAM J. Sci. Stat. Comput., (1991).
[96] Ioannis Pantazopoulos and Spyros Tzafestas, *Occlusion culling algorithms: A comprehensive survey*, Journal of Intelligent and Robotic Systems, 35 (2002), pp. 123–156.
[97] M. Passenbrunner, *Pointwise estimates for b-spline gram matrix inverses*, Journal of Contemporary Mathematical Analysis, 49 (2014), pp. 78–90.
[98] O Penrose and JL Lebowitz, *On the exponential decay of correlation functions*, Commun. Math. Phys., 184 (1974).
[99] Juan C. Pichel, Juan a. Lorenzo, Francisco F. Rivera, Dora B. Heras, and Tomás F. Pena, *Using sampled information: Is it enough for the sparse matrix-vector product locality optimization?*, Concurr. Comput. Pract. Exp., 26 (2014), pp. 98–117.
[100] Ali Pinar and Michael T Heath, *Improving performance of sparse matrix-vector multiplication*, Proc. 1999 ACM/IEEE Conf. Supercomput. - Supercomput. '99, (1999), pp. 30–es.
[101] Hanan Samet, *The design and analysis of spatial data structures*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990.
[102] ———, *Foundations of Multidimensional and Metric Data Structures*, Morgan Kaufmann, 2006.
[103] Rahul S Sampath, Santi S Adavani, Hari Sundar, Ilya Lashuk, and George Biros, *Dendro: parallel algorithms for multigrid and amr methods on 2: 1 balanced octrees*, in Proceedings of the 2008 ACM/IEEE conference on Supercomputing, IEEE Press, 2008, p. 18.
[104] Rahul S Sampath, Hari Sundar, and Shravan K Veerapaneni, *Parallel fast gauss transform*, in Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, IEEE Computer Society, 2010, pp. 1–10.
[105] Tamas Sarlos, *Improved Approximation Algorithms for Large Matrices via Random*, In Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), (2006).
[106] Scott A Sarra, *Regularized symmetric positive definite matrix factorizations for linear systems arising from rbf interpolation and differentiation*, Engineering Analysis with Boundary Elements, 44 (2014), pp. 76–86.
[107] Robert Schaback, *Convergence of unsymmetric kernel-based meshless collocation methods*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 333–351.
[108] Stefan Schamberger and JM Wierum, *A Locality Preserving Graph Ordering Approach for Implicit Partitioning: Graph-Filing Curves.*, ISCA PDCS, (2004).
[109] K.L. Schuchardt, B.T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T.L. Windus, *Basis set exchange: A community database for computational sciences*, J. Chem. Inf. Model., 47 (2007), pp. 1045–1052.
[110] Varun Shankar, Grady B Wright, Robert M Kirby, and Aaron L Fogelson, *A radial basis function (rbf)-finite difference (fd) method for diffusion and reaction–diffusion equations on surfaces*, Journal of scientific computing, 63, pp. 745–768.
[111] Volker Strassen, *Gaussian elimination is not optimal*, Numer. Math., 13 (1969), pp. 354–356.
[112] Matthias Teschner, Bruno Hiedelberger, Matthias Müller, Danat Pomeranets, and Markus Gross, *Optimized Spatial Hashing for Collision Detection of Deformable Objects*, Vmv 2003, (2003), p. 8.
[113] Sivan Toledo, *Improving memory-system performance of sparse matrix-vector multiplication*, IBM J. Res. Dev., 41 (1997), pp. 711–726.

[114] AI Tolstykh and DA Shirobokov, *On using radial basis functions in a finite difference mode with applications to elasticity problems*, Computational Mechanics, 33 (2003), pp. 68–79.
[115] Christopher Umans, *Group-theoretic algorithms for matrix multiplication*, in Proceedings of the 2006 International Symposium on Symbolic and Algebraic Computation, ISSAC '06, New York, NY, USA, 2006, ACM, pp. 5–5.
[116] David van der Spoel, Aldert R. van Buuren, Emile Apol, Pieter J. Meulenhoff, D. Peter Tieleman, Alfons L. T. M. Sijbers, Berk Hess, K. Anton Feenstra, Erik Lindahl, Rudi van Drunen, and Herman J. C. Berendsen, *Gromacs User Manual version 3.0*, Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: http://www.gromacs.org, 2001.
[117] J. Voit, *The Statistical Mechanics of Financial Markets*, Theoretical and Mathematical Physics, Springer Berlin Heidelberg, 2006.
[118] Michael S. Warren, *2HOT: an improved parallel hashed oct-tree n-body algorithm for cosmological simulation*, in Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis, ACM, 2013, p. 72.
[119] Michael S. Warren and John K. Salmon, *A parallel hashed Oct-Tree N-body algorithm*, in Proceedings of the 1993 ACM/IEEE conference on Supercomputing, Supercomputing '93, New York, NY, USA, 1993, ACM, pp. 12–21.
[120] ———, *A parallel, portable and versatile treecode*, SIAM, Philadelphia, 1995, ch. 1.
[121] ———, *A portable parallel particle program*, Computer Physics Communications, 87 (1995), pp. 266 – 290.
[122] ———, *Simulating the motion of 322,000,000 self-gravitating particles*, 1997. ACM Gordon Bell Prize.
[123] Mac Will, *Bcsstk14: Bcs structural engineering matrices (linear equations) roof of the omni coliseum, atlanta.* `http://math.nist.gov/MatrixMarket/data/Harwell-Boeing/bcsstruc2/bcsstk14.html`, 1985.
[124] D Wise, *Costs of quadtree representation of nondense matrices*, J. Parallel Distrib. Comput., 9 (1990), pp. 282–296.
[125] David S. Wise, *Representing matrices as quadtrees for parallel processors: extended abstract*, SIGSAM Bull., 18 (1984), pp. 24–25.
[126] ———, *Ahnentafel Indexing into Morton-Ordered Arrays, or Matrix Locality for Free*, in Euro-Par 2000 Parallel Processing, Arndt Bode, Thomas Ludwig, Wolfgang Karl, and Roland Wismüller, eds., vol. 1900 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2000, pp. 774–783.
[127] David P Woodruff, *Sketching as a Tool for Numerical Linear Algebra* , 10 (2015), pp. 1–139.
[128] Da Yan, James Cheng, Yi Lu, and Wilfred Ng, *Blogel: A block-centric framework for distributed computation on real-world graphs*, Proc. VLDB Endow., 7 (2014), pp. 1981–1992.

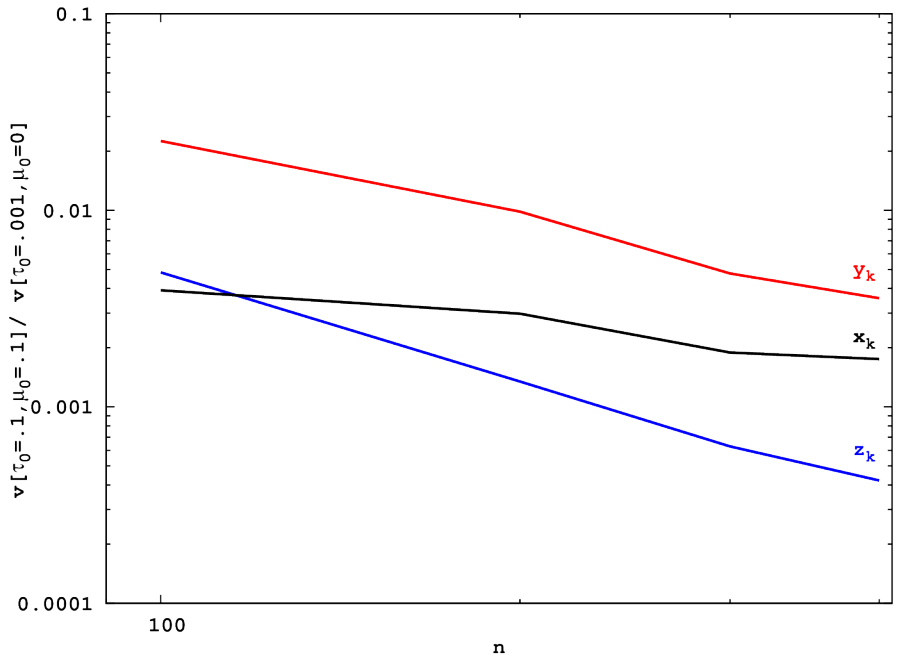

FIG. 7.4. *Complexity reduction in metric square root iteration for the periodic 6-311G** water sequence. Shown is the ratio of lensed product volumes for the regularized MAYEBOO approximation with respect to the unregularized (MAYSS) approximation.*

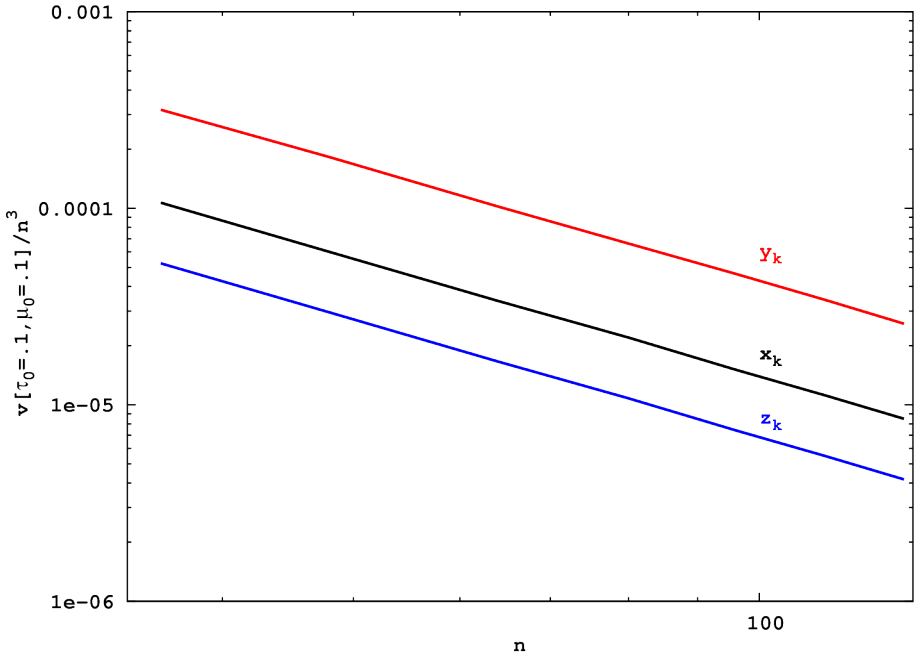

FIG. 7.5. *Complexity reduction in square root iteration for the* $\kappa(\boldsymbol{s}) = 10^(10)$ *sequence. Shown is the ratio of lensed product volumes for the regularized MAYEBOO approximation with respect to the unregularized MAYSS approximation, which we take to be* n^3.